REPORT NO. P66-135
HAC REF. NO. A7747

THIRD QUARTERLY REPORT
CONTRACT NO. NAS 5-9637

PARAMETRIC ANALYSIS OF MICROWAVE AND LASER SYSTEMS FOR COMMUNICATION AND TRACKING

JUNE 1966

AEROSPACE GROUP

HUGHES

HUGHES AIRCRAFT COMPANY
CULVER CITY CALIFORNIA

U. S. Government Agencies Only

FACILITY FORM 602
N67-32165
(ACCESSION NUMBER)
387
(PAGES)
CR-77943
(NASA CR OR TMX OR AD NUMBER)
1
(THRU)
07
(CODE)
(CATEGORY)

Report No. P66-135
HAC Ref. No. A7747

Third Quarterly Report

for

PARAMETRIC ANALYSIS OF MICROWAVE AND LASER SYSTEMS FOR COMMUNICATION AND TRACKING

(6 March 1966 - 6 June 1966)

Contract No. NAS 5-9637

Prepared by
Aerospace Group
Hughes Aircraft Company
Culver City, California

L. S. Stokes
Program Manager

K. L. Brinkman
Associate Program Manager

for

Goddard Space Flight Center
Greenbelt, Maryland

U. S. Government Agencies Only

SUMMARY

PROGRAM STATUS

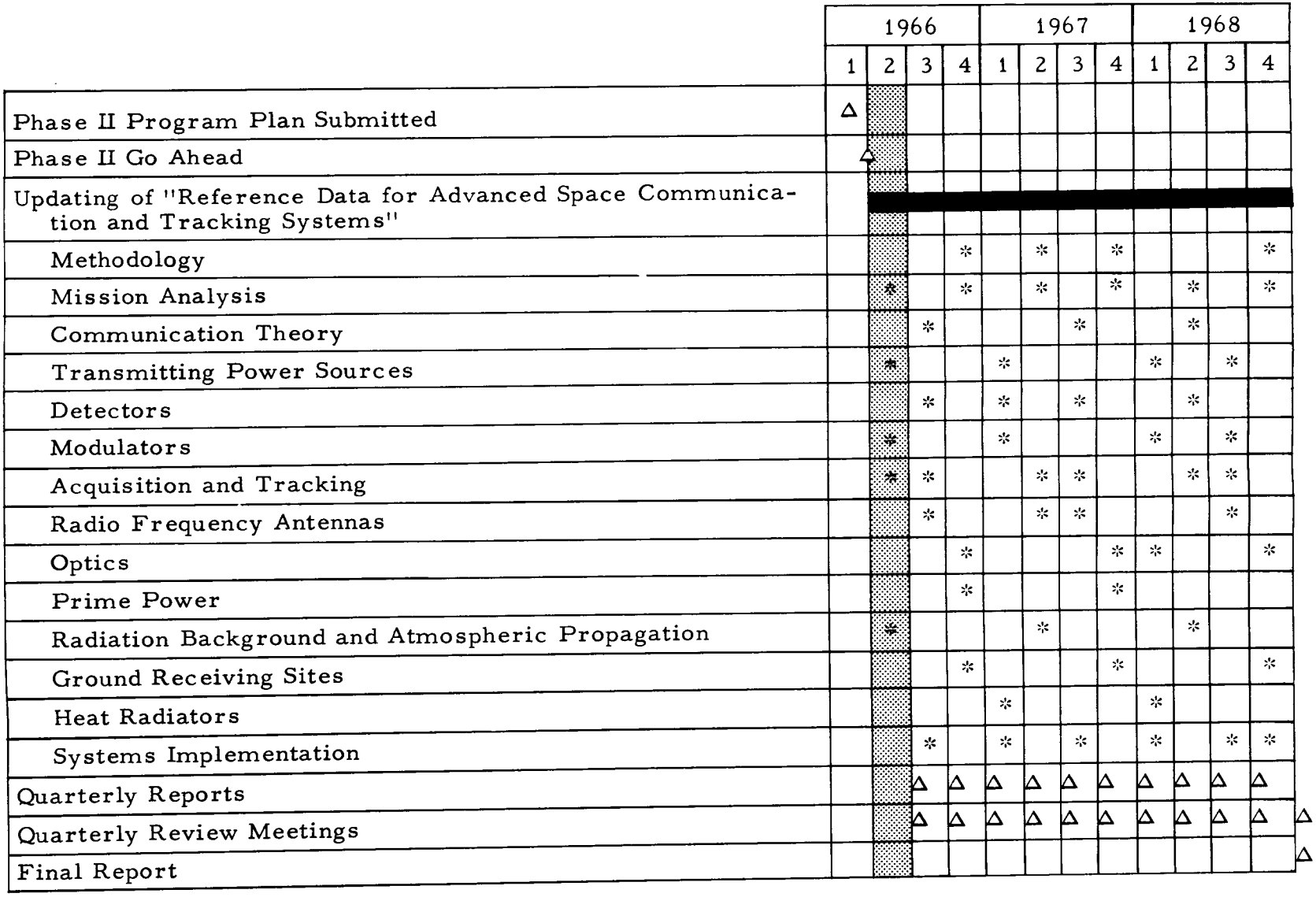
The National Aeronautics and Space Administration, in its space exploration program, has promoted space communications through hardware development and technical studies. August 6, 1965, the Goddard Space Flight Center, NASA, awarded the Hughes Aircraft Company a study contract for Parametric Analysis of Microwave and Laser Systems for Communication and Tracking, NAS 5-9637.

This study contract has three basic purposes: first, to collect the information and conclusions of previous studies and present them in a readily accessible form. Secondly, to formulate a "Reference Data for Advanced Space Communications and Tracking Systems" which will contain: (1) a methodology for solving space communication and tracking problems, (2) parametric studies of the parameters involved in the methodology and, (3) a state-of-the-art documentation of the parameter values. Finally, the third major purpose of the study is to evaluate the capability and amenability to modification of the available world wide communication and tracking system to the increased performance requirements of future microwave and optical communications systems.

The Parametric Analysis of Microwave and Laser Systems for Communication and Tracking is to be conducted in two phases. Phase I had a six-month duration and Phase II has a thirty-three month duration. The study is presently at the three-month point of Phase II. The program plan for Phase II is shown in Figure 1. To a large degree this plan indicates periodic updating and enlargement of the first issue of "Reference Data for Advanced Space Communication and Tracking Systems."

The purpose of the Reference Book is to enable a designer to determine an optimum communication and tracking system for a deep space vehicle. In order to determine the optimum system, a great deal of basic information must be accumulated and properly organized for a variety of topics over a wide range of possible transmitting frequencies. It is not the purpose of this compilation of data and methodology to favor

PRECEDING PAGE BLANK NOT FILMED.



* Technical section updated

Figure 1. Program plan for parametric analysis of microwave and laser systems for communications and tracking.

any particular frequency but to allow the designer to look at the merits of each configuration and to use the one that best matches his constraints and requirements.

The effort during this quarter and the effort of the next quarter has and will be directed toward enlarging and updating the Phase II Program Plan submitted 6 February 1966. The last quarter of 1966 will utilize this compilation of data and the updating of the methodology section, to be completed during that quarter, to formulate a document that should have a wide application to Space Communication and Tracking Systems. The Methodology has been developed to a limited extent during this last quarter. This development has resulted in a center-line design method which determines the optimum split of invested effort among the several key parameters (e.g., spacecraft antenna size, transmitter power, ground antenna size, etc.) of a Communication and Tracking System. This design has been implemented in a relatively straightforward computer program and has produced results, for rather simple problems, which are intuitively satisfying.

SUMMARY BY TECHNICAL SECTIONS

This issue of "Reference Data for Advanced Space Communications and Tracking Systems" is summarized by section in the following material. The Reference data updating includes: Missions Analysis, Transmitting Power Sources, Modulators, Acquisition and Tracking, and Background Radiation and Atmospheric Attenuation. Before a brief summary of the material in each of these sections is given, a summary of a general space communications section will be presented. This section, which immediately follows this summary, considers the overall communications problem and asks the question "at what frequency can a given data rate be achieved with the least weight?" The approach taken is to document the explicit and implicit frequency relationships of the transmission equation to determine the best operating frequencies. This is done from the standpoint of the highest product of signal-to-noise ratio (with a constant transmitted power) and data bandwidth. The analysis indicated three frequency bands shown as peaks in Figure 0-4. The weight required

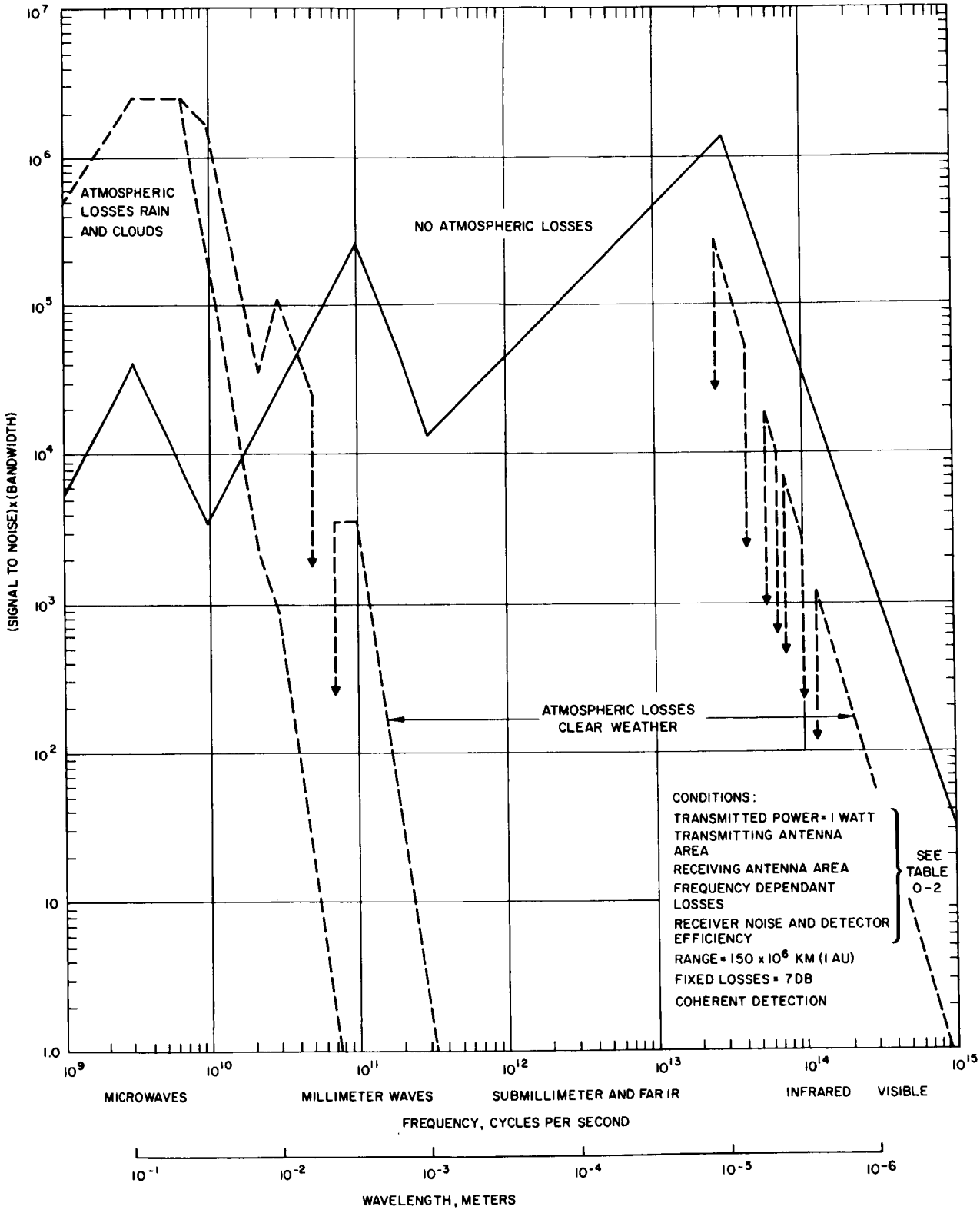


Figure 0-4. Communication system performance.

to build a typical communications system in these three bands and at 0.5μ were calculated. The results are given in Table 0-6. Here it is seen that the 2.3 G system requires the least weight.

Summary of Section 3.0, Missions Analysis

Missions Analysis, Section 3.0, summarizes manned and unmanned space missions envisioned over the next several decades. The primary objectives of each mission are described and implied communication requirements are discussed.

The choice of a space communication system for a particular mission must take into account

1. Probable objectives of the mission under consideration
2. Reflection of these mission objectives into communication system requirements.

This involves definition of communication range and system lifetime requirements and their effects on data transmission rate and data processing and storage facility requirements. More basic limitations are the restrictions on communication system weight, volume, power, aperture size and pointing accuracy imposed by technological limitations. Relevant trade-offs between the required communication rate and data storage and processing requirements are considered. The present and projected states of the data storage and processing arts are briefly reviewed as a guide to this trade-off. Communication ranges and mission durations characteristic of the anticipated missions are presented, as are spacecraft weight limitations imposed by launch vehicles.

Summary of Section 5.0, Transmitting Power Sources

Transmitting Power Sources, has been significantly enlarged in the area of solid state power sources in the microwave region and in the area of CO_2 laser technology in the optical region.

Solid state microwave sources such as IMPATT oscillators and Gunn effect oscillators hold promise of producing significantly higher RF powers than have been available with other solid state devices. The

Total-Power Required	2.3 Ghz 773 watts	94 Ghz 2000 watts	0.5 μ 5300 watts	10.6 μ 800 watts
Solar Panel (80% eff) 0.125 lbs/watt	121 lbs	313 lbs	830 lbs	125 lbs
Regulator-Batteries 0.025 lb/watt-0.05 lb/watt-hr	97 lbs	250 lbs	663 lbs	100 lbs
Transmitter				
2.3 Ghz—0.037 lbs/watt 0.5 μ—25 lbs/watt 94 Ghz—0.04 lbs/watt 10.6 μ—3.3 lbs/watt	10 lbs	16 lbs	132 lbs	400 lbs
Heat exchanger 0.07 lb/watt	35 lbs	112 lbs	370 lbs	48 lbs
Antenna/Optics System*	40 lbs	38 lbs	4 lbs	220 lbs
Tracker/Control System	105 lbs	105 lbs	9 lbs	105 lbs
0.5 μ—2.2 lb/lb optics	---	---	---	---
others—105 lbs	408 lbs	834 lbs.	2008 lbs	998 lbs

* 2.3/94 Ghz—2.5 lbs/ft. includes servo and mounting structure
0.5 μ—0.625 lbs/dia^2 (inches) includes servos and mounting structure
10.6 μ—0.17 lbs/dia^2 (inches) body-fixed optics

Table 0-6. Spacecraft transmitting system weight.

theory and performance of these devices has been included in this issue. Figure 5-11 illustrate the performance that has been obtained with Gann oscillators.

The CO_2 laser, with its efficiently produced energy has received a great deal of attention in recent months. This issue of Transmitting, Power Sources, documents the theory for the CO_2 laser, discusses its scaling laws, modes of operation and gives the state-of-the-art in CO_2 lasers. Figure 5-38 is typical of the parametric variations of the CO_2 laser while Table 5-8 lists current state-of-the-art in CO_2 lasers.

Transmitting power sources for space communications requires more than power output. The power may be required to be at a single frequency and be stable at that frequency. For these reasons methods of laser mode locking and methods of laser stabilization have been given. These discussions point out several methods of achieving the desired results. However the methods implemented to date are quite complicated and are available only in a laboratory environment.

Summary of Section 8.0, Acquisition and Tracking

The efforts in Acquisition and Tracking in the last quarter have been directed towards documentation of the basic constraints within which an acquisition and tracking system must operate and in examining the effects of these constraints on some general configurations.

Although only a limited amount of quantitative data is presently available, the perturbative effects of the atmosphere on both the transmitted and received beams are discussed. The anomolous effects of turbulence and other inhomogenities are documented. They produce, for example, beam steering errors which may be as large as ±15 μrad in the presence of strong turbulence.

Ranges of probable angular errors are discussed. Included are the minimum pointing errors which can be predicted by the DSIF. Also given are the errors associated with predicting the rotation of the line-of-sight vector as illustrated in Figure 8-6.

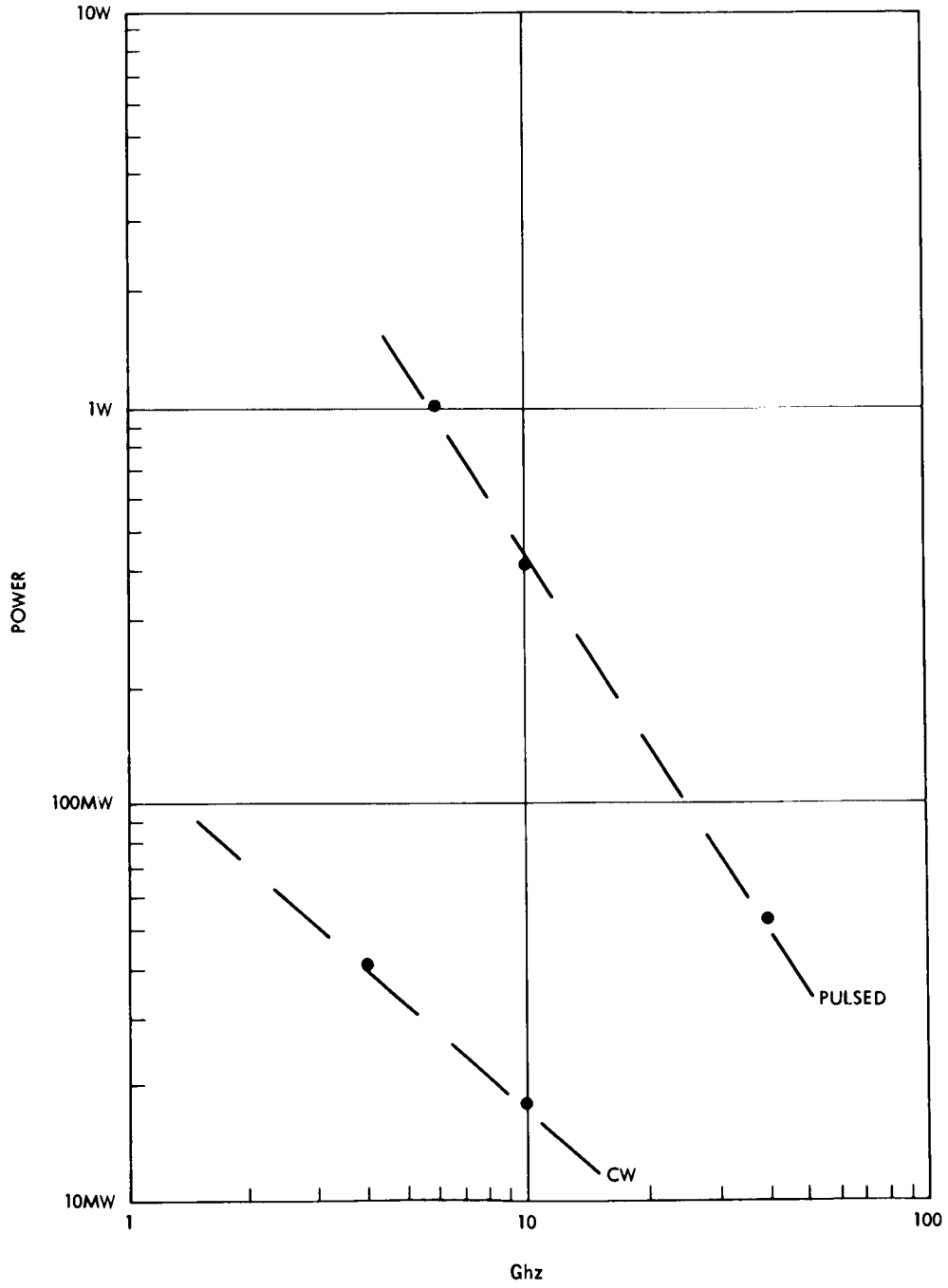


Figure 5-11. Gunn oscillator performance.

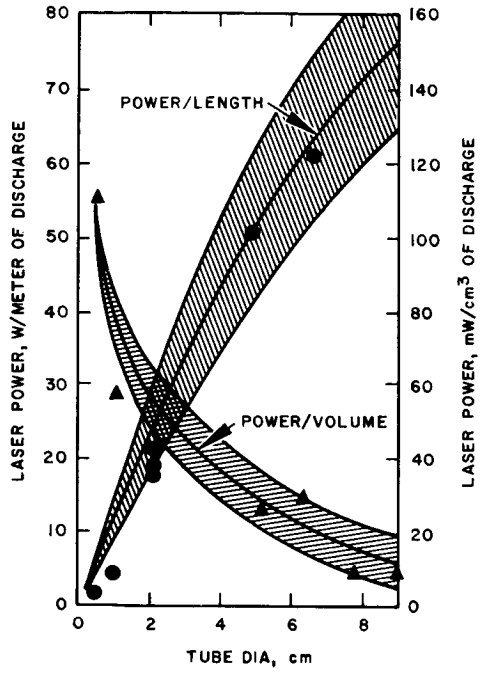


Figure 5-38. Variation of laser power per unit volume of discharge and laser power per unit length of discharge with discharge diameter.

	P_{out} (w)	Length (m)	Diam (cm)	Eff. (%)	Fig. of Merit		Comments	Organization
					P/λ^2 (w/m^2)	$\eta^2 P/L\lambda^2$ (w/m^3)		
1)	500	10	5	15.5	4.5×10^{12}	1.1×10^{10}	Multimode output Flowing gas system	Raytheon
2)	10	0.5	1	10	8.9×10^{10}	1.8×10^{9}	Single TEM_{oo} mode output Sealed-off tube	HRL
3)	150	2.4	6	18	1.4×10^{12}	1.8×10^{10}	Multimode output Flowing gas system	CGE (France)
4)	130	4	7.5	13	1.2×10^{12}	5×10^{9}	Multimode output Flowing gas system	BTL

Table 5-8. State-of-the-art CO_2 lasers (May 1966).

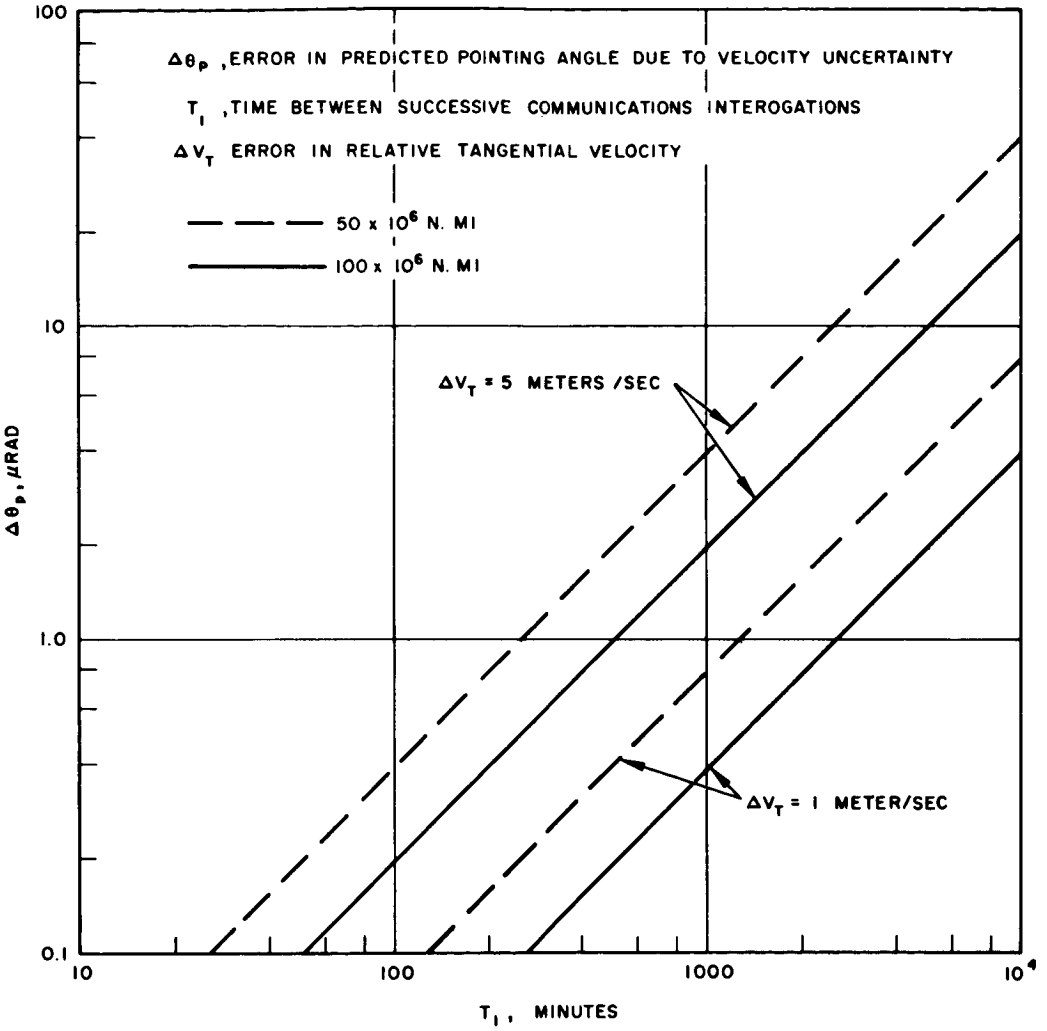


Figure 8-6. Error in predicted pointing angle time between successive versus communications interrogation.

The state-of-the-art for star and sun sensors was documented to a limited degree. Table 8-3, repeated here, is representative of current development for these sensors. The field of view provided by these trackers is adequate to cover the predictive and systematic errors associated with initial acquisition and with tracking.

Summary of Section 7.0, Optical Modulators

In the present revision, emphasis is placed on the more important advances in optical modulation technology which are currently in progress. A section on Elasto-Optic Modulation has been added. This is of timely importance because of the urgent need to develop techniques for modulating the high-power sources being made available at 10.6 microns by the CO_2 laser. The subject of internal modulation of laser oscillators has been considerably expanded. This technology, similarly to that of acoustic (elasto-optic) modulation, is still in an early stage of development. Its importance lies in its promise to effect great economies in size, weight, and driving power required as compared with existing optical modulation devices which are external to the laser oscillator.

The section on Electro-Optic Modulation has been revised in several ways to improve and clarify the presentation. Data has been included on lithium niobate ($LiNbO_3$), which has recently become available with sufficient size and optical quality to show promise at least for internal modulation. This crystal has good optical transmission from about 4000 Å to beyond 4 microns. Available data on lithium tantalate, another promising new crystal, also has been added.

The possibility of placing some initial approximate values on the modulator performance/burden relationships defined in Section 2.0, Methodology, was considered. In our judgment it is still premature to make meaningful assessments of the weight, cost, or driving power required as a function of information bandwidth. These values will vary with the particular optical source considered, and may not be continuously variable as a function of information bandwidth. At 10.6 microns, a wavelength of great current interest for space applications, there is essentially no existing technology for wideband modulation of high-power

Instrument	Performance Data	Burden Parameters
ITT Federal Labs OAO Boresighted Star Tracker	1. Star Magnitude Sensitivity +6 2. FOV 10 Arc Min 3. RMS Accuracy 1.5 Arc Sec 4. Gimballed (Electronically) ±1.5 deg 5. Two-Axis Capability	1. Weight 25 lbs 2. Power Requirements ±28V dc 7.7 watts 3. Size 3" (dia) × 15" Electronics: 5" × 11" × 12"
ITT Federal Labs Dual Mode Star Tracker	1. Star Magnitude Sensitivity +3 2. FOV 8^{o} × 8^{o} Acquisit. 32' × 32' Track 3. Accuracy (RMS) ±5 Arc Sec 4. Not Gimballed 5. Two-Axis Capability	1. Weight 9.5 lbs 2. Power 8 watts (26V dc) 3. Size 5" × 10.5" × 5"
ITT Federal Labs Electro-Optical Sensing Head for Star Tracking	1. Sensitivity +2.5 2. FOV 1^{o} × 1^{o} 3. Accuracy (RMS) ±9 Arc Sec 4. Not Gimballed 5. Two-Axis Capability	

Table 8-3. Star trackers.

Instrument	Performance Data	Burden Parameters
Kollsman Inst. Co. Gimballed Seeker	1. Sensitivity +2 2. FOV $1^{o} \times 1^{o}$ 3. RMS Accuracy ± 5 Arc Sec 4. Gimballed (Mechanical) $\pm 4.3^{o}$ 5. Two-Axis Capability	1. Weight 42.5 lbs 2. Power 29 watts 3. Size N. A.
Bendix Corp. Star Sensor	1. Sensitivity >+2.1 2. FOV $10^{o} \times 10^{o}$ 3. Accuracy $< 0.1^{o}$ 4. Not Gimballed 5. Single Axis	1. Weight 5.0 lbs 2. Power 8 watts 3. Size 4" × 7" × 10"
ITT Federal Labs Sensor Design Proposed for SERT-III	1. Sensitivity >2.1 2. FOV $10^{o} \times 10^{o}$ 3. Accuracy $< 0.012^{o}$ 4. Single Axis	1. Weight 8.0 lbs 2. Size 2 1/4" (dia) × 8.5"

Table 8-3 (continued). Star trackers.

optical beams. A speculative approach is to use acoustic techniques applied internally to the laser oscillator. There is, however, no valid basis at this time for making a prediction of the performance/burden relationships which might result from development of this technique.

A summary of the state-of-the-art in the optical modulator field is given in Table 7-4 which lists characteristics of some electro-optic modulators as of June 1966.

Summary of Section 12.0, Background Radiation and Atmospheric Propagation

External noise is of great significance in determining ultimate communication system performance. Section 12 presents such information as is currently available concerning external noise sources relevant to optical and microwave space communication systems. The principal types of noise encountered are:

1. Cosmic noise, originating outside the solar system
2. Solar system noise, other than terrestrial, originating from the sun, the planets, or their satellites
3. Terrestrial noise, originating from the earth or its surrounding atmosphere.

Propagation of optical radiation through the terrestrial atmosphere including the effects of atmospheric turbulence in introducing phase and polarization changes, image dancing, beam scanning, and scintillation is considered and its implication with regard to system performance is discussed. The effect of atmospheric turbulence in introducing phase and amplitude fluctuations in radio frequency radiation is briefly discussed.

Parameter	TM_{oo} Mode Cavity Modulator	TW Strip-Line Intensity Modulator	Zig-Zag Intensity Modulator	Strip Line Polarization Modulator (NASA Contract)	Multi-Element Intensity Modulator	Resonant SSBSC Modulator	TW Strip-Line SSB Modulator	TW Strip-Line Intensity Modulator
Development Status	Built at BTL	Built at Sylvania	Dev. at BTL	Built at Hughes Aircraft Company	Built at Hughes Aircraft Company	Built at Hughes Aircraft Company	Built at Hughes Aircraft Company	Designed at Sylvania
Material	KDP	KDP	CuCl	KDP	KDP	KDP	KDP	KDA
Crystal Dimensions	1 cm long × 2.5mm dia.	100 cm long × 2mm wide	6.7 cm long × 1mm sq	50 × 0.4 × 0.4 cm	16 each 1/4" × 1/4" × 1/2"	2 each 1 1/2" long × 1/2" dia.	80 × 0.4 × 0.4 cm	18 × 0.2 × 0.2 cm
Optical Attenuation	0.1 dB	6 dB	2.6 dB	1.5 dB	0.3 dB	1.0 dB	2.0 dB	Unknown
Useful Optical Range	0.4 - 1.5 μ	0.4 - 1.5 μ	0.4 - 20 μ	0.4 - 1.5 μ	0.4 - 1.5 μ	0.4 - 1.5 μ	0.4 - 1.5 μ	0.4 - 1.5 μ
Modulation Frequency	3 Gc	Baseband	Baseband	Baseband	Baseband	850 Mc	200 Mc	Baseband
3 dB Bandwidth	4 Mc	3 Gc	> 10 Gc	30 Mc*	10 Mc*	5 Mc	100 Mc	3 Gc
Modulating Power	1.5 W	12 W	1.6 W	20 W	12 W	3 W	10 W	5 W
Modulation Index	0.13	~1.0	0.5	0.5	0.5	0.01	0.3	0.3
Extinction Ratio	Unknown	Unknown	Unknown	8:1	250:1	NA	NA	Unknown
Weight	Unknown	Unknown	Unknown	20 lbs	2 lbs	15 lbs	25 lbs	Unknown
Size	Unknown	Unknown	Unknown	144 in^3	30 in^3	850 in^3	216 in^3	Unknown

*3dB Bandwidth limited by modulator driver and not by the electro-optic structure.

Table 7-4. Characteristics of some electro-optic modulators - June 1966.

CONTENTS

SUMMARY iii

SYSTEM PERFORMANCE 0-1

REFERENCE DATA FOR ADVANCED SPACE COMMUNICATIONS AND TRACKING

1.0 INTRODUCTION 1-1

3.0 MISSION ANALYSIS 3-1

5.0 TRANSMITTING POWER SOURCES 5-1

7.0 OPTICAL MODULATORS 7-1

8.0 ACQUISITION AND TRACKING 8-1

12.0 BACKGROUND RADIATION AND ATMOSPHERIC CONTINUATION 12-1

0.0 SYSTEM PERFORMANCE

The potential telecommunication performance for a deep space communication link (from a space vehicle to earth) is considered in this section as a function of frequency. Consideration is given to both the basic and practical limitations which apply.

First, as a measure of telecommunication system performance, the product of signal quality and quantity is derived from the one-way transmission equation, and the explicit frequency-dependent terms are separated from the system parameters. Next, link performance is investigated as a function of the explicit frequency dependency and as it is influenced by the implicit restrictive frequency dependencies imposed on the "constant" parameters of the transmission link equation. Included are the losses due to atmospheric absorption and the influence of atmospheric distortion on receiver aperture size. Total link performance is then given as a product of the various frequency dependent parameters. The optimum operating frequencies indicated as a result of this analysis will depend on the relative values taken as parametric restrictions. The analysis indicates three frequency bands which promise higher performance than the frequencies which are immediately adjacent. A detailed weight breakdown has been made for a communication system operating in each of these bands and at one other wave length of interest, 0.5 μ. Finally, consideration is given to operational system requirements and their effect on the overall system implementation. This consideration includes the relative role of earth receiving stations versus satellite receiving stations.

0.1 PERFORMANCE CRITERIA

The performance of a telecommunication link may be measured against many criteria. Chief among these criteria are amount of data, trustworthiness of the data, and economy in transmission of this data. In this study, primary emphasis is placed on achieving an improvement in the amount of data while maintaining its quality. Economy of data

transmission remains important but is placed in a secondary role at this time. For in general, realistic cost evaluations must be related to specific candidate systems as they develop from basic systems studies and research. With the product of transmitted data quantity and quality as a performance criterion the one way transmission equation is written below where the signal-to-noise ratio, S/N, represents signal quality, the bandwidth, B, represents signal quantity, and other symbols are as usually defined:

$$\left(\frac{S}{N}\right)(B) = \frac{G_T G_R P_T \lambda^2 L_o}{R^2 (4\pi)^2 hf \left[1 + e^{hf/kT} - 1)^{-1}\right]} \tag{1}$$

Insertion of the functional frequency dependencies of the parameters of this equation (e.g., $G_R = 4\pi A_R / \lambda^2$) gives:

$$\left(\frac{S}{N}\right)(B) = \frac{L}{R^2 hc^2} A_T A_R P_T L_f M \tag{2}$$

where L_f represents the frequency dependent losses and M, an ultimate figure of merit, is given by[1]

$$M = \frac{f}{1 + (e^{hf/kT} - 1)^{-1}} \tag{3}$$

Equations (2) and (3) document the explicit functional frequency dependencies and separate them from the "constants" of the transmission link equation. However, some of these "constants" are indirectly dependent on frequency in terms of practical system considerations. Consider, for example, the frequency dependency placed on a spaceborne transmitting antenna due to a pointing accuracy restriction. Such a restriction sets a minimum beamwidth (or maximum gain), thereby requiring the diameter of the transmitting antenna to decrease inversely as frequency increases. This and other such restrictions must be

included when applying Equation (2) to the process of optimum frequency selection. The explicit functional frequency dependency given by the figure of merit, M, is discussed below as are the implicit frequency dependencies due to the various practical limitations imposed on the transmitting antenna area, A_T; the receiving antenna area, A_R; the transmitted power, P_T; and losses due to atmospheric transmission represented by L_f.

Functional Frequency Dependency, M

Equation (3) compares ultimate system performance for communication systems operating at different frequencies but with constant values of antenna area for both transmit and receive, constant power transmitted, and constant losses in transmission. This equation is plotted in Figure 0-1 and indicates that with the above parameters fixed, the performance increases linearly with frequency where quantum noise is dominant and with the square of frequency where the thermal noise is dominant. As indicated in this figure, there is an ultimate system improvement possible of six orders of magnitude in going from present operating frequencies in the microwave band to the visible light spectrum. While the actual achievement of this million-fold improvement in system performance is inhibited by the practical parametric restrictions to be discussed, still the figure of merit, M; does represent an ultimate performance which could be approached upon the investment of suitable effort. It sets a goal, the attainment of which will require the development of new concepts and techniques.

Transmitting Antenna Area

The effective area of a spacecraft transmitting antenna may be limited by one of three practical considerations:

- weight and size limitations imposed by the vehicle,
- antenna fabrication tolerances, and
- transmitting beamwidth demanded by the limitation on pointing accuracy.

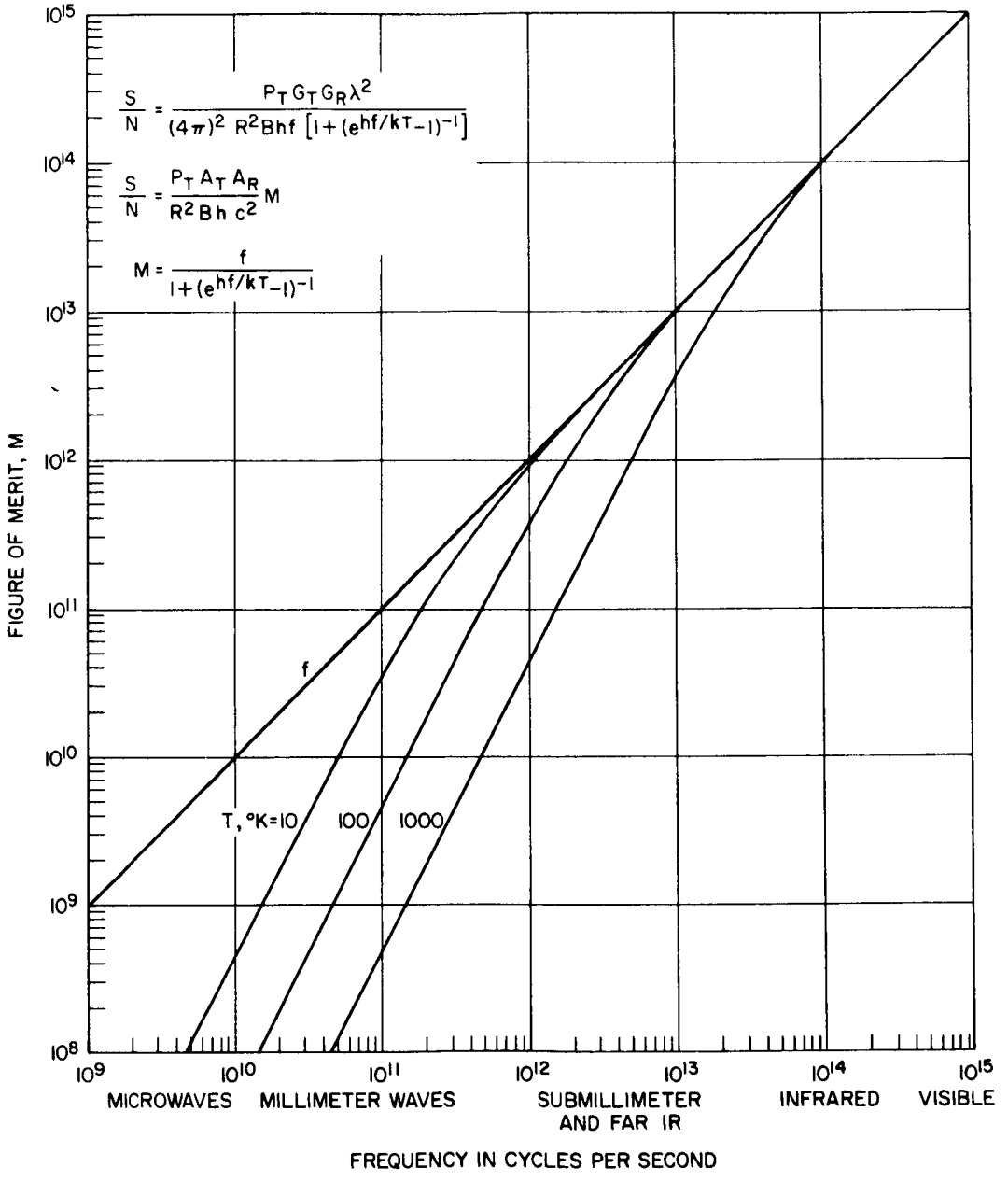


Figure 0-1. Ultimate system figure of merit vs. frequency.

Although there is some interrelation among these factors (for example, an unfurlable antenna may be used for increased diameter at the expense of dimensional tolerances) one will generally dominate within a certain frequency range.

In Figure 0-2 transmitting antenna area is plotted against frequency for different values of the three restrictive conditions. Horizontal lines indicating diameter restrictions could be due to a variety of constraints such as booster characteristics or vehicle weight allocations. Fabrication tolerances are proportional, to a first approximation, to overall dimensions and hence are expressed in normalized form as the ratio of the rms dimensional deviation to the diameter, σ/D. Since a given ratio implies a particular gain limit[2], useful antenna area will be inversely proportional to the square of the frequency for a given σ/D value. Similarly, a given pointing accuracy requires a certain beamwidth and hence represents a gain limit. Thus, the same family of curves can represent the restriction on transmitting antenna area due to either fabrication tolerance or pointing accuracy limitations. The curves, evaluated for an approximate 3 db loss, result in lines of constant negative slope on the log-log plot of Figure 0-2 and are labeled in terms of both σ/D and beamwidth.

With reference then to Figure 0-2, the frequency dependency of the transmitting antenna area may be determined based on an appropriate set of assumptions as to vehicle constraints, practical fabrication tolerances, and achievable pointing accuracy.

Receiving Antenna Area

The same type of restrictions apply to the receiving antenna area as to the transmitting antenna. Thus, the useful area of a single element may be limited by allowable size, achievable tolerances, or (for coherent reception) achievable pointing accuracy. For a satellite receiving station, limiting values will not differ too much from those for a deep space vehicle. On the other hand, for a ground station, receiver diameter will in general not be restricted by installation requirements but for a single element will be subject to a limitation due to atmospheric

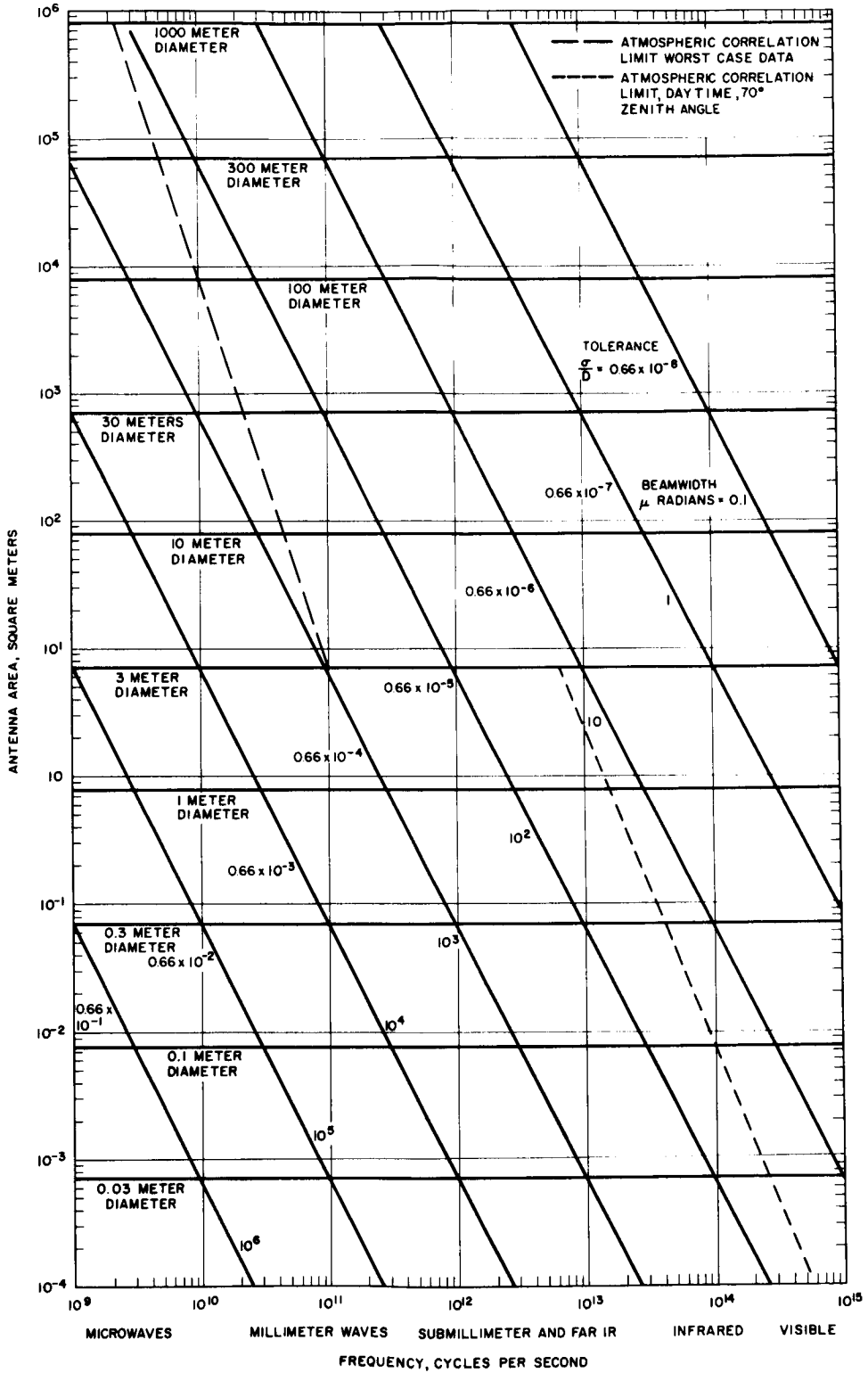


Figure 0-2. Parameter restrictions on antenna area.

distortion of the signal wavefront over the aperture. As discussed in Section 12.0, little quantitative information is available on such atmospheric anomalies as wavefront corrugation and tilt, or unequal illumination of the aperture. However, an indication of the restriction imposed on antenna diameter can be gained based on data available from the NBS experiment[3] at Maui, Hawaii (1956). The assumption is made that the wavefront distortions, corresponding to the phase deviations experimentally measured as a function of baseline will have an effect similar to that caused by dimensional deviations of the antenna surface due to fabrication tolerances. It is further assumed the linear dimensions of the distortion are to a first approximation independent of frequency[4]. A curve based on the worst data from the Maui experiment is plotted with the other restrictions in Figure 0-2.

In the optical region of the spectrum an atmospheric correlation length has been defined which is found[5] to vary as $\lambda^{6/5}$. The effective optical aperture area corresponding to this correlation length for typical daytime conditions of atmospheric turbulence is given in Figure 0-2. This restriction applies only to coherent reception.

Transmitted Power

Frequency dependent restrictions on transmitter power are related to the method of power generation. At microwave frequencies traveling wave tubes have been found most satisfactory in providing high powers, and this capability extends to the millimeter wavelengths. In general, however, maximum power generation capability in TWT amplifiers is inversely proportional to frequency.

At optical frequencies the principal limitation on power generated is set by heat dissipation in the laser cavity as a result of low conversion efficiencies. There is no significant systematic dependency on frequency, performance being related to the different specific laser mechanisms.

In either portion of the spectrum an increase in power can always be achieved by paralleling units. Hence the effective limitation is set by weight, size, and power supply restrictions.

Atmospheric Losses

Transmission losses occur in the atmosphere due to absorption of electromagnetic radiation by the constituent gases (especially water vapor) and in the optical region, due to aerosol and molecular scattering. Absorption occurs in bands in the spectrum rather than as a systematic variation; this spectral absorption is indicated in Section 12.0. Scattering from aerosols generally accounts for the predominant loss at low altitudes in the visible region, where the transmission is roughly proportional to wavelength.

Rain and fog begin to have a serious effect on transmission in the millimeter region of the RF spectrum, with the effect due to fog increasing at optical frequencies. It is evident that an optical communication link would be effectively blacked out by cloud cover. Table 0-1[6] shows the percentage of time that various sites experience obscuration by clouds and indicates the seriousness of this problem.

Receiver Noise

Ideal and projected system noise performance is given in Figure 0-3[1]. Essentially the noise has no explicit frequency dependency other than that already included in the figure of merit, M. It depends in practice, however, on external noise sources and on achievable equivalent noise temperatures in the RF region and detector quantum efficiencies in the optical region. Presently achievable performances in the RF and projected performance in the optical spectrum are indicated in the figure. Although a receiver noise figure was not explicitly included in the one-way transmission equation, it has the effect of reducing the value of the figure of merit, M. The noise figure and external noise (see Section 2) can be accounted for in the RF region by taking the equivalent rather than actual input temperature as the effective value of M. In the optical region M should be reduced by the effective quantum efficiency of the detector when considering the performance to be expected of a real system.

Station Coordinates (Long.)	Lat.		Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
New Mexico		1962	36	25	27	15	5	12	49	32	48	19	26	22
253°27'(E)	+32°25'	1963	34	27	21	29	30	13	48	57	28	16	27	17
South Africa		1962	42	29	31	26	4	6	1	5	8	29	53	37
028°15'(E)	-25°58'	1963	50	23	27	28	15	21	20	2	7	30	55	49
Australia		1962	32	21	21	9	34	12	22	22	27	38	15	18
136°46'(E)	-31°06'	1963	16	22	22	22	51	35	47	21	18	17	18	22
Spain		1962	52	32	81	43	33	15	21	15	34	47	40	43
353°48'(E)	+36°28'	1963	72	49	39	39	35	32	4	19	22	19	60	45
Peru		1962	81	71	56	38	10	2	5	6	35	22	42	55
288°30'(E)	-16°28'	1963	88	95	63	56	23	2	6	15	21	26	13	52
Iran		1962	28	44	33	54	10	1	18	16	03	nil	14	38
052°31'(E)	+29°38'	1963	17	40	30	47	30	7	15	23	1	12	35	22
Curacao		1962	52	43	56	53	74	66	46	38	60	39	45	50
291°10'(E)	+12°05'	1963	70	59	51	68	70	40	64	55	63	50	65	62
Florida		1962	55	40	62	55	36	61	41	55	52	45	55	43
279°53'(E)	+27°01'	1963	57	58	47	28	50	42	44	29	49	32	49	43
Argentina		1962	53	52	38	72	51	27	39	32	22	29	35	43
294°54'(E)	-31°57'	1963	*	35	40	27	39	33	36	32	57	46	46	30
Hawaii		1962	23	41	56	39	33	5	38	18	20	16	26	33
203°45'(E)	+20°43'	1963	59	17	70	78	52	28	29	23	39	39	32	30

*No photography attempted for a 3 week period when the mirror was removed for realuminizing.

Table 0-1. Baker Nunn sites showing percent of viewing time lost due to clouds.

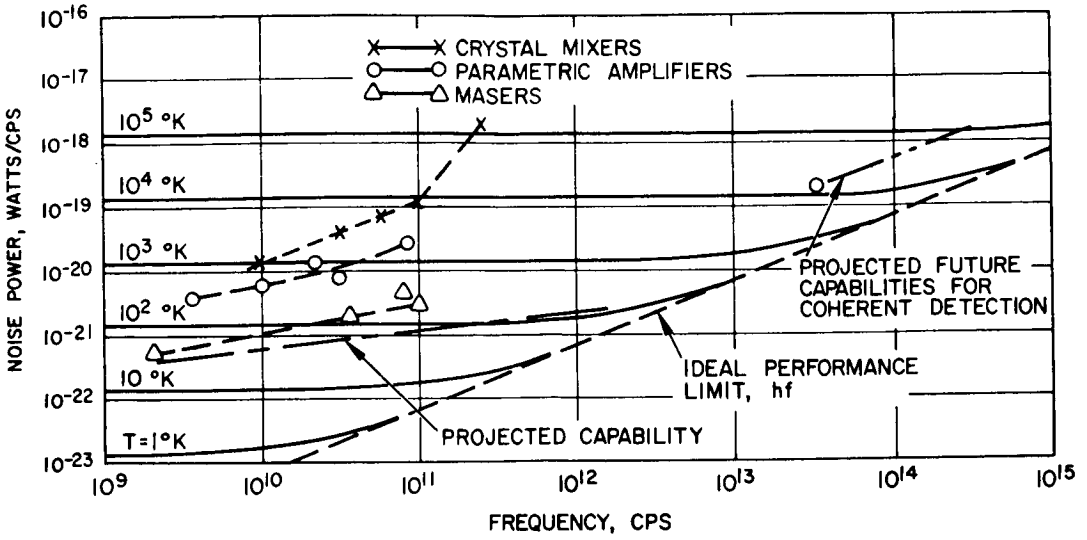


Figure 0-3. System noise performance.

0.2 SAMPLE PERFORMANCE CALCULATION

System performance versus frequency may now be calculated based on the one-way transmission equation and on various limiting values of the parameters as appropriate. To serve as an example and give an indication of the sort of curve that results, a calculation is made taking representative parametric restrictions as tabulated in Table 0-2. Practical restrictions imposed by the atmosphere are considered for three cases of interest:

a. No atmospheric losses (corresponding to a satellite receiving station),
b. Atmospheric losses due to clear weather conditions, and
c. Atmospheric losses due to poor weather conditions.

The specific parameter values for the several frequencies at which changes occur in the chosen parametric restrictions are given in Table 0-3. The resultant product of signal quality (S/N) and signal quantity (B) is also tabulated.

The product (S/N) (B) is plotted in Figure 0-4 for the three cases considered with reference to the atmosphere. The sharp peaks and

Frequency Dependent Parameter	No Atmospheric Losses (Satellite Receiving Station)	Atmospheric Losses, Clear Weather	Atmospheric Losses, Including Rain and Fog
Power transmitted, P_T	Power taken as one watt for all frequencies		
Transmitting antenna area, A_T limits (see Figure 0-3)	10 meter maximum diameter to frequency of 3×10^9 cps; $\sigma/D = 0.66 \times 10^{-3}$ to frequency of 10^{10} cps; 3 meter maximum diameter to 10^{11} cps; $\sigma/D = 0.66 \times 10^{-4}$ to 3×10^{11} cps; 1 meter maximum diameter to 3×10^{13} cps; beamwidth of 10 μ radian to 10^{15} cps.		
Receiving antenna area, A_R limits (see Figure 0-3)	Same as A_T limits above	100 meter diameter to 10^{10} cps; atmospheric correlation limit to 10^{11} cps; 3 meter diameter to 7×10^{12} cps; atmospheric correlation limit to 10^{15} cps.	
Atmospheric frequency dependent losses, L_f	None	Losses per Figures 12-34, 12-37, and 12-38	Losses per Figures 12-36 and 12-40
Receiver noise and detector efficiency	Noise temperature taken as 10°K to 2×10^{11} cps. Quantum efficiency taken as 0.5 to 10^{15} cps.	Noise temperature per Figure 12-17 (10 degree elevation angle) to 5×10^{10} cps 300 degrees to 5×10^{12} cps. Quantum efficiency 0.5 to 10^{15} cps.	Noise temperature per Figure 12-17 (10 degree elevation angle) to 10^{10} cps 300 degrees to 5×10^{12} cps. Quantum efficiency 0.5 to 10^{15} cps.

Table 0-2. Sample frequency dependent parameter limitations.

Frequency, cps	L/R^2hc^2, watts^{-1} meters^{-4}	P_T, watts	A_T, meters^2	A_R, meters^2	L_f	M, cps	(S/N)(B), cps
10^9	1.49×10^{-7}	1	78	78	1.0	6×10^6	5.4×10^3
3×10^9			78	78		4.5×10^7	4.1×10^4
10^{10}			7.1	7.1		4.5×10^8	3.4×10^3
2.2×10^{10}			7.1	7.1		2×10^9	1.5×10^4
3×10^{10}			7.1	7.1		4×10^9	3.0×10^4
10^{11}			7.1	7.1		3.5×10^{10}	2.6×10^5
3×10^{11}			0.78	0.78		1.5×10^{11}	1.4×10^4
3×10^{13}			0.78	0.78		1.5×10^{13}	1.4×10^6
3×10^{14}			7.1×10^{-3}	7.1×10^{-3}		1.5×10^{14}	1.1×10^3
10^{15}	1.49×10^{-7}	1	7.1×10^{-4}	7.1×10^{-4}	1.0	5×10^{14}	3.7×10

Table 0-3a. System parameter values, no atmosphere.

Frequency, cps	L/R^2hc^2, watts^{-1} meters^{-4}	P_T, watts	A_T, meters^2	A_R, meters^2	L_f	M, cps	(S/N)(B), cps
10^9	1.49×10^{-7}	1	78	7.8×10^3	0.96	6×10^6	5.2×10^5
3×10^9			78	7.8×10^3	0.96	3×10^7	2.6×10^6
10^{10}			7.1	7.8×10^3	0.94	2.2×10^8	1.7×10^6
2.2×10^{10}			7.1	6.8×10^2	0.42	1.2×10^8	3.6×10^4
3×10^{10}			7.1	2.7×10^2	0.76	5×10^8	1.1×10^5
10^{11}			7.1	7.1	0.32	1.5×10^9	3.6×10^3
3×10^{11}			0.78	7.1	$\sim 10^{-4}$	1.5×10^{10}	1.3
3×10^{13}			0.78	0.15	0.6	1.5×10^{13}	1.6×10^5
3×10^{14}			7.1×10^{-3}	5.3×10^{-4}	0.6	1.5×10^{14}	5×10
10^{15}	1.49×10^{-7}	1	7.1×10^{-4}	2.8×10^{-5}	—	5×10^{14}	—

Table 0-3b. System parameter values, clear weather.

Frequency, cps	L/R^2hc^2, watts^{-1} meters^{-4}	P_T, watts	A_T, meters^2	A_R, meters^2	L_f	M, cps	(S/N)(B), cps
10^9	1.49×10^{-7}	1	78	7.8×10^3	0.96	6×10^6	5.2×10^5
3×10^9			78	7.8×10^3	0.96	3×10^7	2.6×10^6
10^{10}			7.1	7.8×10^3	0.5	7×10^7	2.9×10^5
2.2×10^{10}			7.1	6.8×10^2	0.04	8×10^7	2.3×10^3
3×10^{10}			7.1	2.7×10^2	0.02	1.5×10^8	8.6×10^2
10^{11}			7.1	7.1	$\sim 10^{-5}$	1.5×10^9	0.1
3×10^{11}			0.78	7.1	–	1.5×10^{10}	–
3×10^{13}			0.78	0.15	–	1.5×10^{13}	–
3×10^{14}			7.1×10^{-3}	5.3×10^{-4}	–	1.5×10^{14}	–
10^{15}	1.49×10^{-7}	1	7.1×10^{-4}	2.8×10^{-5}	–	5×10^{14}	–

Table 0-3c. System parameter values, rain and fog.

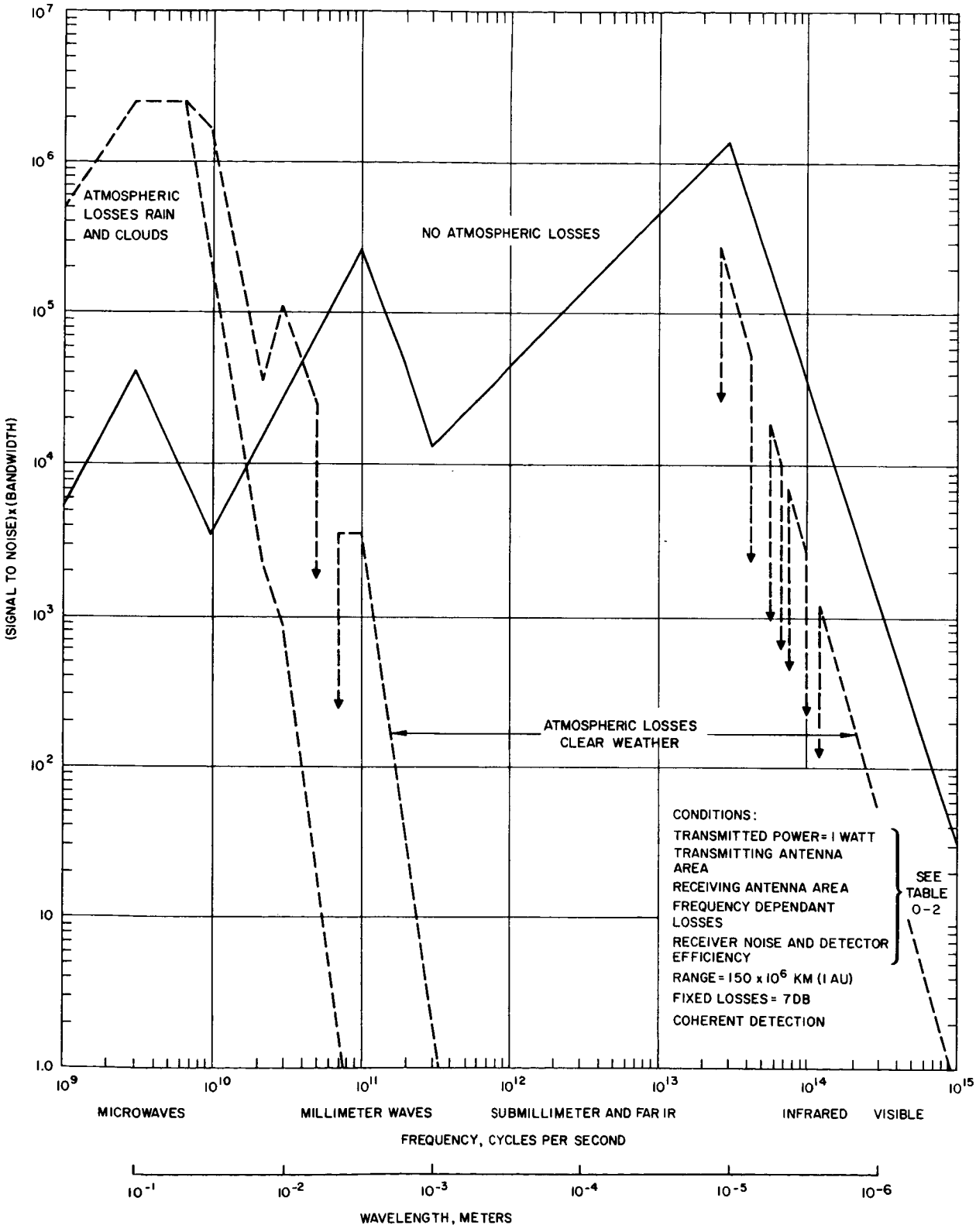


Figure 0-4. Communication system performance.

valleys result from the discontinuous nature of the parametric restrictions selected, and their exact positions are dependent on the particular values taken as representative of practical limits.

The curve for a satellite receiving station shows three maxima. These occur at 3 Ghz, 100 Ghz, and 3×10^{13} cps, or a wavelength of approximately 10 microns, and are directly related to the maximum frequencies for which constant antenna sizes were taken for both the transmitter and receiver. The cutoffs after the first two peaks occur due to the more stringent fabrication tolerances required at shorter wavelengths, while the cutoff in the optical region results from a limitation in pointing accuracy.

Similar maxima are indicated for an earth based receiving site under clear weather conditions except that the millimeter-wave maximum is heavily attenuated and shifted by the atmospheric losses. The optical portion of this curve is reduced slightly from the corresponding satellite receiving station curve by the atmospheric restriction on lateral coherence over the wavefront.

When the restrictions imposed by clouds and rain are considered the familiar single maximum at microwave frequencies results for the product (S/N) (B).

The curves of Figure 0-4 are normalized in three senses: a range of one astronomical unit is taken, a transmitted power of 1 watt is used, and a single-aperture ground antenna is assumed. Thus, increased performance at this range is available by increasing the transmitted power or by increasing the number of ground antennas to form an array.

It should also be noted that the curves in the optical region represent performance limitations on coherent detection. For non-coherent detection, pointing accuracy and atmospheric distortions do not impose restrictions on receiver aperture area so that the reduction in performance with frequency in the optical region (for no atmosphere or for clear weather) will be as the inverse of the frequency rather than the inverse cube of frequency as indicated in Figure 0-4.

Comparison Analysis, Required Power

Figure 0-4 shows three maxima in the (S/N) B versus frequency curve. These occur at 3 Ghz, 100 Ghz, and at a wavelength of approximately 10 microns and correspond to the parameter values chosen.

The potential telecommunication performance for a deep space communication link (from space vehicle to earth) is considered at these three frequency bands plus an incoherent optical system at 0.5 micron. The comparison criteria are transmitted power and total transmitting system weight for a deep space vehicle. In addition, the transmitting power was evaluated for three sets of antennas at 2.3 Ghz. They are: (1) a 16-foot transmitting antenna and a 210-foot receiving antenna, (2) a 32-foot transmitting antenna and a 210-foot receiving antenna, and (3) a 16-foot transmitting antenna and a 744 by 744-foot square receiving antenna.

The parameters involved in minimizing transmitted power may be seen by rearranging Equation 1.

$$P_T = \frac{(S/N)\,(4\pi R/\lambda)^2\,\{hf\,[1 + (1/e^{hf/KT} - 1)]\,B\}}{G_t G_r L_T}$$

The microwave system (2.3 Ghz) is compared to the millimeter system (94 Ghz) in Table 0-4 while the infrared system (10.6 micron) is compared to an incoherent optical system (0.5 micron) in Table 0-5. All systems with the exception of the latter employ coherent detection. Each system is assumed to provide the same error rate, 10^{-5}, for a bit rate of 2×10^6 bits per second at a range of 100 million nautical miles.

Explanatory notes for the parameter values used in Tables 0-4 and 0-5 are given below.

Signal-to-Noise - For an error rate of 10^{-5} the signal-to-noise power ratio is taken as ten for all systems.

Losses - The atmospheric loss was determined for a zenith angle (angle off the vertical) of 75 degrees. (See Section 12.) Antenna efficiencies are included in the antenna gain terms.

Noise Power - Sky temperatures were determined assuming a zenith angle of 75 degrees. For a coherent system the minimum value for noise power is hf B while for incoherent detection the noise is 2 hf B. Noise bandwidth is assumed equal to one-half the bit rate.

Detector Efficiency - The detector quantum efficiency is assumed to be 50 percent. At present, this figure is optimistic; however it is felt that detectors can be made with this quantum efficiency throughout the visible and infrared.

Transmitter Antenna Gain - Several organizations engaged in the development of stowable antenna techniques report that a 30- to 50-foot diameter reflector antenna could be assembled in space to operate at 2 to 3 Ghz with efficiencies of 50 percent and gains of 45 to 50 db. A 94 Ghz, a 15-foot cassegrain parabolic reflector antenna with a gain of 70 db is operating at the Aerospace Corporation.

For the optical frequencies, the antenna size is limited in this calculation by the ability to point a narrow beam. For both 0.5 micron and 10.6 micron the beam size is limited to 10 microradians (2 arc/seconds). This requires a pointing accuracy of about 0.2 arc/second.

Receiver Antenna Gain - A ground based 210-foot antenna is operational in the DSIF. Large (600- to 1000-foot) antennas are being considered for deep space at S band. Reference three points out that an array (10 elements) of 210-foot antennas is economically feasible for very deep space systems.[4] The 10.6μ antenna size is limited by atmospheric correlation. (Figure 0-2.) For 0.5μ, the antenna size is not limited by atmospheric correlation since an incoherent detector is used. The 30.5-foot antenna is proposed here.

	2.3 Ghz		94 Ghz	
Signal-to-Noise, S/N, for error rate of 10^{-5}	10	10 db	10	10 db
Beam Spreading Loss $(4\pi R/\lambda)^2$	λ = 13 cm R = 1.85×10^8Km	-265.1 db	λ = 0.32 cm R = 1.85×10^8Km	-297.2 db
Antenna efficiencies	0.54	—	0.54	—
Total line loss	0.8	-1.0 db	0.7	-1.5 db
Atmospheric Loss	0.95	-0.2 db	0.5	-3.0 db
System Noise Temperature, KT	10^o sky + 25^o receiver	-213.2 db/cps	200^o sky + 200^o receiver	-205.7 db/cps
Noise Bandwidth, B	10^6	60 db/cps		60 db/cps
Noise Power, KTB		-153.2 dbW		-145.7 dbW
Transmitter antenna gain, G_T (beam width)	16 feet (~20 mrad.)	38.8 db	15 feet (~ 1 mrad.)	70 db
Transmitter antenna gain, G_T'	32 feet	44.8 db	—	—
Receiver antenna gain, G_r	210 feet	60 db	15 feet	70 db
Receiver antenna gain, G_r'	744 feet × 744 feet	76 db	—	—
Transmitter Power, P_T, for G_T and G_r	270 watts	24.3 dbW	400 watts	26 dbW
Space Craft Power	773 watts	28.9 dbW	2000 watts	33 dbW
Transmitter efficiency	0.35		0.2	
Transmitter Power, P_T', for G_T' and G_r	67.5W	20.3 dbW		
Transmitter Power, P_T'', for G_T and G_r'	6.8W	8.3 dbW		

Table 0-4. Micro/millimeter transmission from a deep space vehicle to earth.

	0.5 μ incoherent		10.6 μ coherent	
Signal-to-Noise, S/N, for error rate of 10^{-5}	10	10 db	10	10 db
Beam Spreading Loss $(4\pi R/\lambda)^{\lambda}$	$R = 1.85 \times 10^{8}$Km	-373.6 db	$R = 1.85 \times 10^{8}$Km	-347.1 db
Transmitter optics loss	0.8	-1 db	0.7	-1.5 db
Receiver optics loss	0.3	-5.2 db	0.3	-5.2 db
Detector Quantum Efficiency	0.5	-3 db	0.5	-3 db
Atmospheric Loss	0.45	-3.4 db	0.6	-2.2 db
Noise Bandwidth, B	10^{6} hz	60 db/cps	10^{6} hz	60 db/cps
Noise Power hf B (incoherent) 2hf B	8×10^{-13}	-121 dbW	1.9×10^{-14}	-137.2 dbW
Transmitter antenna gain, G_T beam width	2.5 inches 10 μrad	112 db	3 feet 10 μrad	109.4 db
Receiver antenna gain, G_r	30.5 feet	156 db	1.5 feet	101.6 db
Transmitter Power, P_T	5.3 watts	7.2 dbW	120 watts	20.8 dbW
Space Craft Power [XMTR efficiency]	5,300 watts 0.001	37.2 dbW	800 watts 0.15	29 dbW

Table 0-5. Optical/infrared transmission from a space vehicle to earth.

The purpose of Tables 0-4 and 0-5 is to calculate the required transmitted and input power at each frequency considered. These powers are then interpreted in terms of the weight required to provide such a power and these weights are added to the estimated weights of the other communication components to obtain an overall system comparison on the basis of weight. This comparison is tabulated in Table 0-6 for the four different frequencies and in Table 0-7 for the 3 sets of antenna size at 2.3 G cps. As is seen from these tables, the configuration at 2.3 G cps is considerably more attractive than the other frequencies on the basis of weight.

Comparison Analysis, Required

Space Craft Transmitting System Weight

In the weight comparison of the four systems shown in Table 0-6, the transmitting system includes primary power sources, regulator and battery, and heat exchanger which are primarily dependent on total power with regard to weight. The weights of the remaining portion of the system namely antenna/optics and associated tracker/control systems are dependent on antenna/optics diameters. The weight of the optics in the infrared system (10.6 μ) is reasonably low because lightweight body-fixed optics are assumed. Also the tracker is assumed to be a control system type identical to that in the millimeter (94 Ghz) system. If body-fixed optics were not assumed the weight of the optics, including servos and supporting structure would be about 800 pounds while a tracker would go as high as 1800 pounds for a 36-inch telescope. In the 0.5 μ system the optics and tracker are not body-fixed since the telescope diameter is small and their combined weight is less than 1 percent of the total system weight.

As seen in Table 0-8, the systems can be compared on a percentage basis. The optical system (0.5 μ) is almost all total-power dependent (99 percent) while the infrared system (10.6 μ) is more evenly distributed with 32.5 percent total power dependent, while the remaining 67.5 percent is dependent on optics diameter. If body-fixed optics are not employed, the dependence on optics diameter is increased to 81 percent.

Total-Power Required	2.3 Ghz 773 watts	94 Ghz 2000 watts	0.5 μ 5300 watts	10.6 μ 800 watts
Solar Panel (80% eff) 0.125 lbs/watt	121 lbs	313 lbs	830 lbs	125 lbs
Regulator-Batteries 0.025 lb/watt-0.05 lb/watt-hr	97 lbs	250 lbs	663 lbs	100 lbs
Transmitter				
2.3 Ghz–0.037 lbs/watt 0.5 μ–25 lbs/watt 94 Ghz–0.04 lbs/watt 10.6 μ–3.3 lbs/watt	10 lbs	16 lbs	132 lbs	400 lbs
Heat exchanger 0.07 lb/watt	35 lbs	112 lbs	370 lbs	48 lbs
Antenna/Optics System*	40 lbs	38 lbs	4 lbs	220 lbs
Tracker/Control System	105 lbs	105 lbs	9 lbs	105 lbs
0.5 μ–2.2 lb/lb optics	---	---	---	---
others–105 lbs	408 lbs	834 lbs.	2008 lbs	998 lbs

* 2.3/94 Ghz–2.5 lbs/ft. includes servo and mounting structure
0.5 μ–0.625 lbs/dia^2 (inches) includes servos and mounting structure
10.6 μ–0.17 lbs/dia^2 (inches) body-fixed optics

Table 0-6. Spacecraft transmitting system weight.

Antenna Size (Ground) Antenna Size (Transmitter) 2.3 Ghz	210 ft. 16 ft.	210 ft. 32 ft.	744×744 ft. 16 ft.
Total Power Required	773 watts	193 watts	19.4 watts
Solar Panel-Power Regulator-Batteries	218 lbs.	55 lbs.	6 lbs.
Transmitter	10 lbs.	3 lbs.	1 lb.*
Heat exchange	35 lbs.	9 lbs.	1 lb.
Antenna System	40 lbs.	80 lbs.	40 lbs.
Tracker/Control System	105 lbs.	105 lbs.	105 lbs.
	408 lbs.	252 lbs.	153 lbs.

* Transmitters are available from 1 watt to 20 watts weighing one pound in the range of 1-10 Ghz. reference

Table 0-7. Effect of an increase in antenna gain on weight.

System Weight Dependence	2.3 Ghz	94 Ghz	0.5 μ	10.6 μ
Total-Power	64.4%	82.6%	99%	67.5%
Antenna/Optics size	35.6%	17.4%	1%	32.5%*

Table 0-8. Distribution of weight between power and antenna/optics size.

0.3 SYSTEM CONFIGURATIONS

As indicated in the previous section, improved performance may be achieved by an increase in parameter values or by a suitable selection of operating frequency. A major hindrance to the improvement promised at higher frequencies is presented by the attenuating and distorting characteristics of the atmosphere. Space missions which are likely to require the wide signal bandwidths considered in this study are those involving real-time transmission of data and hence a continuous communications capability. Such a requirement can be expected for a flyby mission, a landing mission, or a mission requiring voice communication. At microwave frequencies continuous coverage can be achieved as is seen from Figure 0-4. However, at optical frequencies clouds may cut off a ground site completely. Thus, in order to meet the requirements of a continuous optical link from the spacecraft to the earth within an acceptable atmospheric loss, alternate transmission configurations must be considered.

Optical Transmission Through the Atmosphere

If a pure optical system is to be used with no backup radio link, the only alternative available to provide continuous coverage is multiple ground stations. These stations must be distributed such that at least one is always visible to the space probe for a given phase of the earth. The number of such stations would probably be 5 to 10 times the number of stations operating at a frequency not limited by atmosphere.

In addition to increased costs a second unfortunate consequence of multiple ground stations results from the very narrow beams appropriate to optical communications. These beams will, in general, illuminate only a portion of the earth—an area of perhaps 1000 miles or less in diameter. Thus, if it is necessary to change the ground receiving station due to atmospheric problems, it will also be necessary to redirect the transmitted beam of the spacecraft to the new earth site. This redirection must be accomplished by an "open loop" command from earth with its accompanying time delays.

Radio Transmission Through the Atmosphere

Figure 0-4 indicates that a radio frequency may be selected for which the concommitant attenuation will be acceptable; and indeed, data such as shown in this figure combined with the limitations imposed by radio noise sources have led to the selection of the present frequencies used for deep space communications. However, if the ultimate potential of Figure 0-1 is to be exploited, a higher frequency must be used. This suggests a hybrid system, one which uses a suitable microwave frequency for propagation through the atmosphere and which then employs a much higher frequency, in the millimeter or optical region, for transmission between the deep space probe and a relay satellite.

Summary

Possible configurations for future deep space communication systems that can provide the desired improvement in performance over present systems can be summarized as follows:

- A direct space vehicle to earth link may be employed, operating in the lower microwave region from about 2 to 7 Ghz. This frequency will permit transmission through the atmosphere with acceptably low atmospheric attenuation and low noise contribution from the sun, stars, etc. Improved link performance must then be achieved by appropriate improvements in noise performance and transmitted power, and in both transmitting and receiving antenna gains. Increases in gain generally imply larger antenna areas or antenna arrays for the earth receiving station, but some contribution to the gain of the spacecraft antenna may be achieved by an increase in frequency within the 2 to 7 Ghz range.
- A direct probe-to-earth link may be employed operating at a high microwave or optical frequency. The higher operating frequency is selected to obtain greater system performance potential. Multiple receiving sites must be established in

order to assure—with a given probability—that excessive attenuation will not occur because of rain or clouds.

- A hybrid link may be chosen utilizing a high microwave frequency or an optical frequency from the probe to a relay satellite to achieve the high potential of these frequencies. A lower microwave frequency can then be used to relay the data through the atmosphere to (existing) ground stations.

If the higher performance potential of millimeter or optical wavelengths is to be exploited for deep space communication links requiring full-time operational capability, it is evident that either satellite receivers or multiple ground stations are needed. An analysis of the economic implications of these two configurations will be necessary to provide a basis for comparison between them, and for their evaluation with reference to a microwave link. Before valid economic systems analyses can be performed, however, the systems to be analyzed require better definition.

0.4 REFERENCES

1. K. L. Brinkman and L. S. Stokes, Deep Space Communication: System Trade-Off Parameters, IEEE Winter Convention on Aerospace & Electronic Systems, Los Angeles, February 2-4, 1966.

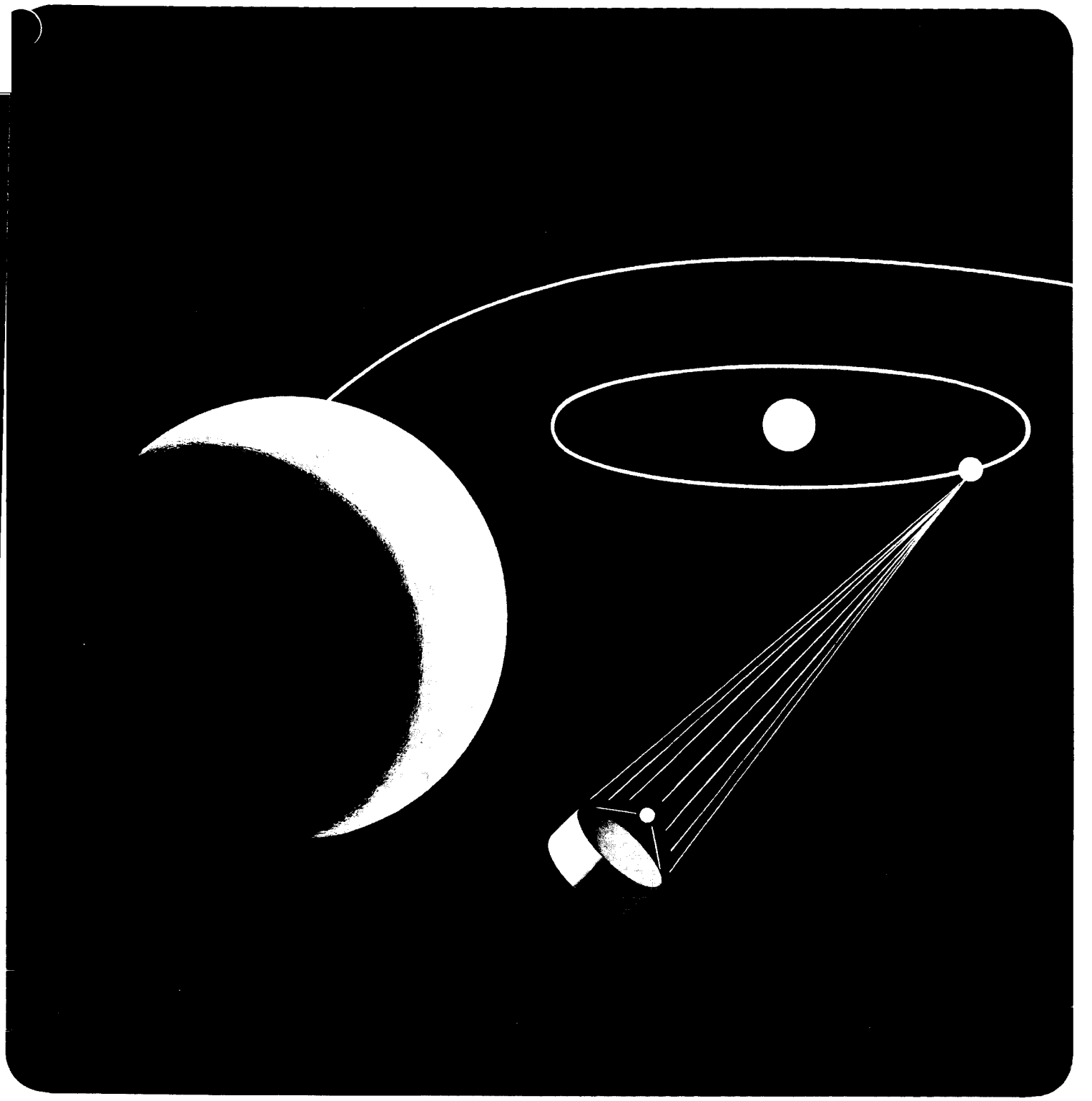
2. John Ruze, Physical Limitations on Antennas, TR No. 248, Research Laboratory of Electronics, MIT, ASTIA/AD 62351, October 1952.

3. K. A. Norton, et al., An Experimental Study of Phase Variations in Line-of-Sight Microwave Transmissions, N. B. S., Monograph 33, November 1961.

4. P. D. Potter, W. D. Merrick, and A. C. Ludwig, Large Antenna Apertures and Arrays for Deep Space Communications, TR No. 32-848, Jet Propulsion Laboratory (NASA), CIT, November 1, 1965.

5. D. L. Fried, The Effect of Wave Front Distortion on the Performance of an Ideal Optical Hetrodyne Receiver and an Ideal Camera, Proc. Conference on Atmospheric Limitations to Optical Propagation, pp 192-241, Boulder, Colorado, 1965.

6. Research in Space Science, Special Report No. 176, Smithsonian Institution Astrophysics Observatory, 17 May 1965.



HUGHES

HUGHES AIRCRAFT COMPANY
CULVER CITY CALIFORNIA

Report No. P66-135

SECOND ISSUE

REFERENCE DATA FOR ADVANCED SPACE COMMUNICATION AND TRACKING SYSTEMS

6 June 1966

Contract No. NAS 5-9637

Prepared by
Aerospace Group
Hughes Aircraft Company
Culver City, California

for
Goddard Space Flight Center
Greenbelt, Maryland

CONTENTS

	Page	Current Issue
1.0 Introduction	1-1	6 June 66
1.1 Purpose	1-1	
1.2 Reference Data Format	1-3	
1.2.1 Introduction	1-3	
1.2.2 Theory	1-3	
1.2.3 Performance	1-3	
1.2.4 Burden Relationships	1-3	
1.2.5 Nomenclature	1-4	
1.2.6 References	1-4	
1.3 Front Matter	1-4	
2.0 Communication Systems Optimization Methodology	2-1	6 Feb 66
2.1 Introduction	2-1	
2.2 Optical Communication Systems Optimization Methodology	2-2	
2.2.1 Optical Communications Components Analysis	2-2	
2.2.2 Optical Communication Theory Analysis	2-26	
2.2.3 Optical Systems Analysis	2-31	
2.2.4 Optical Systems Evaluation	2-34	
2.3 Optical Systems Methodology Example	2-42	
2.3.1 Optical Communications Components Analysis Example	2-42	
2.3.2 Optical Communication Theory Analysis Example	2-45	
2.3.3 Optical Systems Analysis Example	2-48	
2.3.4 Optical Systems Evaluation Example	2-48	
3.0 Mission Analysis	3-1	6 June 66
3.1 Introduction	3-1	
3.2 Probable Mission Objectives	3-1	
3.2.1 Earth Orbital Operations	3-1	

	Page	Current Issue
3.2.2 Lunar Exploration	3-5	6 June 66
3.2.3 Planetary Exploration	3-6	
3.3 Communication System Requirements for Probable Missions	3-8	
3.3.1 Data Transmission Requirements	3-8	
3.3.2 Communication Range	3-19	
3.3.3 Communication System Lifetime Requirements	3-19	
3.3.4 Communication System Weight Restrictions	3-25	
3.4 Conclusion	3-27	
3.5 References	3-28	
4.0 Communication Theory	4-1	6 Feb 66
4.1 Optical Detection	4-2	
4.1.1 Optical Detection Methods	4-2	
4.1.2 Direct Detection	4-3	
4.1.3 Heterodyne Detection	4-7	
4.1.4 Homodyne Detection	4-13	
4.1.5 Optical Detection Statistics	4-16	
4.2 Optical Modulation	4-19	
4.2.1 Optical Modulation Methods	4-19	
4.2.2 Optical Intensity Modulation	4-21	
4.2.3 Optical Frequency Modulation	4-22	
4.2.4 Optical Pulse Position Modulation	4-22	
4.2.5 Optical PCM/Intensity Modulation	4-24	
4.2.6 Optical PCM/Polarization Modulation	4-26	
4.3 Communication Coding	4-31	
4.3.1 Data Compression	4-31	
4.3.2 Data Representation	4-34	
4.3.3 Synchronization	4-35	

	Page	Current Issue
5.0 Transmitting Power Sources	5-1	6 June 66
5.1 Introduction	5-1	
5.2 Radio Frequency Sources	5-1	
5.2.1 Radio Frequency Source Theory	5-2	
5.2.2 Available Source Performance	5-9	
5.2.3 Power-Burden Relationships	5-26	
5.3 Optical Frequency Transmitting Sources	5-29	
5.3.1 Optical Frequency Source Theory	5-29	
5.3.2 Theory of the Argon Ion Laser	5-33	
5.3.3 Characteristics of Mode-Coupled Gas Lasers	5-48	
5.3.4 Laser Stabilization	5-54	
5.3.5 The CO_2 Laser	5-79	
5.3.6 Available Optical Source Performance	5-93	
5.3.7 Power/Burden Relationships	5-104	
5.4 Nomenclature	5-109	
5.5 References	5-113	
6.0 Detectors	6-1	6 Feb 66
6.1 Introduction	6-1	
6.2 Radio Frequency Detectors	6-1	
6.2.1 Micro/Millimeter-Wave Detectors	6-1	
6.2.2 Detectivity/Burden Relationships	6-7	
6.3 Detectors in the Submillimeter Range	6-9	
6.4 Comparison and Evaluation of Detectors	6-10	
6.4.1 Figures of Merit	6-10	
6.4.2 Available Detector Performance	6-10	
6.4.3 Detectivity/Burden Relationships	6-13	
6.4.4 IF Amplifier	6-14	
6.5 Optical Frequency Detectors	6-14	
6.5.1 Optical Frequency Detector Theory	6-14	
6.5.2 Available Detector Performance	6-45	
6.6 References	6-65	

	Page	Current Issue
7.0 Optical Modulators	7-1	6 June 66
7.1 Introduction	7-2	
7.2 Electro-Optic Modulation	7-4	
7.2.1 Theory of Electro-Optic Modulation	7-4	
7.2.2 Design of Intensity Modulators	7-8	
7.2.3 Design of Optical Frequency Translators	7-17	
7.2.4 Design of Phase Modulators	7-29	
7.2.5 Design of Polarization Modulators	7-31	
7.3 Elasto-Optic Modulation	7-32	
7.3.1 Theory of the Elasto-Optic Effect	7-32	
7.3.2 Theory of Ultrasonic Diffraction of Light	7-36	
7.3.3 Modulation Techniques	7-38	
7.3.4 Properties of Ultrasonic Modulators	7-41	
7.4 Internal Modulation	7-47	
7.4.1 Internal Polarization Modulation	7-47	
7.4.2 Other Methods of Internal Modulation	7-49	
7.5 Modulator Performance-Modulator/Burden Relationships	7-52	
7.6 Nomenclature	7-55	
7.7 References	7-58	
8.0 Acquisition and Tracking	8-1	6 June 66
8.1 Introduction	8-1	
8.1.1 Study Status	8-1	
8.1.2 Summary of the Study during the Last Quarter	8-1	
8.1.3 Study Plan for Next Quarter	8-4	
8.2 General Theoretical Considerations	8-5	
8.2.1 Geometrical Considerations	8-5	
8.2.2 Atmospheric Effects	8-6	
8.2.3 Manned versus Unmanned Vehicles	8-8	
8.2.4 Deep-Space Vehicle Station	8-10	

		Page	Current Issue
8.2.5	Earth Acquisition by the DSV	8-12	6 June 66
8.2.6	DSV Tracking and Pointing Loops	8-15	
8.2.7	Stabilization Techniques	8-19	
8.2.8	DSV Acquisition Subsystems	8-24	
8.2.9	DSV Tracking Loop Subsystem	8-27	
8.2.10	Lead Angle and Lead Angle Prediction Errors	8-29	
8.2.11	Pointing Error Budgets	8-33	
8.2.12	Angle Readout and Nonorthogonality Errors	8-36	
8.2.13	Sensor Boresight Errors	8-42	
8.2.14	Earth Tracking Subsystem	8-42	
8.2.15	Pointing Error Budget for an Earth Station	8-44	
8.2.16	Earth Satellite Considerations	8-45	
8.3	Star Trackers and Sensors	8-46	
8.3.1	General	8-46	
8.3.2	Sun Sensors	8-46	
8.4	Acquisition and Tracking/Burden Relationships	8-49	
8.5	Nomenclature	8-49	
8.6	References	8-63	
9.0	Radio Frequency Antennas	9-1	6 Feb 66
9.1	Introduction	9-1	
9.2	Antenna Theory for Large Apertures	9-1	
9.2.1	Introduction	9-1	
9.2.2	Gain Degradation Due to Predictable Systematic (Non-Random) Phase Errors	9-3	
9.2.3	Gain Degradation Due to Random Errors	9-5	
9.3	Available Antenna Types and Performance	9-8	
9.3.1	Reflector-Type Antenna	9-8	
9.3.2	Lens-Type Antenna	9-9	

	Page	Current Issue
9.3.3 Multiple-Beam Antennas	9-12	6 Feb 66
9.3.4 Self-Phased and Adaptive Arrays	9-18	
9.3.5 Phased Array	9-30	
9.3.6 Frequency Scanned Array	9-30	
9.4 Antenna Performance/Burden Relationships	9-30	
9.5 References	9-35	
10.0 Optics	10-1	6 Feb 66
10.1 Introduction	10-1	
10.2 Optical Implementation Choices	10-1	
10.3 Optical Parameters/Burden Relationships	10-8	
10.3.1 Optical Manufacturability	10-8	
10.3.2 Mechanical Manufacturability	10-19	
10.3.3 Alignment	10-21	
10.4 References	10-26	
11.0 Spacecraft Prime Power Generation	11-1	6 Feb 66
11.1 Prime Power Sources for Power Levels Under One Kilowatt	11-1	
11.1.1 Fuel Cell Systems	11-1	
11.1.2 Primary Battery Systems	11-5	
11.1.3 Photovoltaic Array-Battery Power Systems	11-7	
11.1.4 Radioisotope Thermoelectric Generator-Battery Power System	11-14	
11.2 Prime Power Sources for Power Levels Greater than One Kilowatt	11-21	
11.2.1 Solar Power Systems	11-22	
11.2.2 Nuclear Power Systems	11-23	
12.0 Background Radiation and Atmospheric Propagation	12-1	6 June 66
12.1 Introduction	12-1	
12.2 Radio Frequency Background	12-1	
12.2.1 Comparison of Radio Frequency Noise Sources	12-1	

	Page	Current Issue
12.3 Cosmic Background Radiation	12-9	6 June 66
12.3.1 Diffuse Background	12-9	
12.3.2 Radio Stars	12-14	
12.3.3 Hydrogen Line Emission by Interstellar Gas Clouds	12-14	
12.3.4 Solar System Background	12-18	
12.3.5 Terrestrial Atmospheric Radiation	12-25	
12.4 Optical Background	12-26	
12.4.1 Cosmic Background Radiation	12-26	
12.4.2 Solar System Background	12-26	
12.4.3 Terrestrial Atmospheric Background	12-37	
12.5 Propagation of Radio Frequency Radiation through the Terrestrial Atmosphere	12-47	
12.5.1 Attenuation	12-47	
12.5.2 Turbulence Effects	12-47	
12.6 Propagation of Optical Frequency Radiation in the Terrestrial Atmosphere	12-51	
12.6.1 Attenuation	12-51	
12.6.2 Turbulence effects	12-54	
12.6.3 Loss of Spatial Coherence	12-55	
12.6.4 Scintillation	12-59	
12.6.5 Beam Steering	12-61	
12.7 References	12-69	
13.0 Ground Receiving Sites	13-1	6 Feb 66
13.1 Introduction	13-1	
13.2 Optical Reception at the Earth's Surface	13-1	
13.2.1 Seeing	13-1	
13.2.2 Optical Ground Sites	13-5	
13.3 Available Radio System Sites	13-9	
13.3.1 Deep Space Network	13-9	
13.3.2 Unified S-Band System	13-17	
13.3.3 Satellite Tracking and Data Acquisition Network (STADAN)	13-21	
13.4 References	13-31	

	Page	Current Issue
14.0 Heat Rejection System	14-1	6 Feb 66
14.1 Introduction	14-1	
14.2 Pressure Drop	14-1	
14.2.1 Condensing Systems	14-1	
14.2.2 Non-condensing Systems	14-5	
14.3 Heat Transfer	14-6	
14.3.1 Temperature	14-6	
14.3.2 Fin Effectiveness	14-7	
14.3.3 Radiator Area Requirements	14-9	
14.3.4 Emissivity	14-11	
14.4 Radiator Specific Weight	14-11	
14.5 References	14-15	

ILLUSTRATIONS

Figure 2-1	Systems optimization flow chart	2-1
Figure 2-2	Optical communication components analysis flow chart	2-3
Figure 2-3	Optical communication theory analysis flow chart	2-4
Figure 2-4	Optical systems analysis flow chart	2-5
Figure 2-5	Optical systems evaluation flow chart	2-6
Figure 2-6	Field of view and focal length geometric relationship	2-8
Figure 2-7	Receiver focal length vs. receiver field of view	2-9
Figure 2-8	Receiver optical system fabrication cost vs. receiver field of view	2-9
Figure 2-9	Laser power versus wavelength	2-15
Figure 3-1	Orbital missions	3-2
Figure 3-2	Lunar missions	3-2
Figure 3-3	Unmanned planetary missions	3-3
Figure 3-4	Manned planetary missions	3-3
Figure 3-5	Time table of future manned space missions	3-9
Figure 3-6	Television bandwidth requirements	3-16
Figure 3-7	Projected communication requirements for near planets. Message time, or data rate can be improved by technological advances. Data rate appears as bits per second for a typical planetary distance of 240 million miles, with mission capabilities represented by various bit rates	3-17
Figure 3-8	Mercury 1968: time of flight versus launch date	3-20
Figure 3-9	Venus 1970: time of flight versus launch date	3-21
Figure 3-10	Mars 1973: time of flight versus launch date	3-22
Figure 3-11	Jupiter 1973: time of flight versus launch date, Type I	3-22
Figure 3-12	Mercury 1968: Earth-Mercury communication distance versus launch date	3-23
Figure 3-13	Venus 1970: Earth-Venus communication distance versus launch date	3-23
Figure 3-14	Mars 1973: Earth-Mars communication distance versus launch date	3-24

Figure 3-15	Jupiter 1973: Earth-Jupiter communication distance versus launch date, Type I . . .	3-24
Figure 3-16	Launch vehicle capabilities[9] - payloads and destination. The horizontal characteristic velocity scale shows velocity a vehicle needs at 100 n. mi. altitude (not orbit) to reach destination, assuming a minimum energy trajectory	3-26
Figure 4-1	Communication system model	4-1
Figure 4-2	Direct detection model	4-4
Figure 4-3	Local oscillator reference heterodyne detection model	4-8
Figure 4-4	Local oscillator reference homodyne detection model	4-14
Figure 4-5	Difference channel detection method . . .	4-26
Figure 4-6	Probability of detection error as a function of signal and noise level based upon exact analysis and gaussian approximation	4-29
Figure 4-7	Communication coding	4-32
Figure 4-8	Bit representative forms	4-35
Figure 5-1	Power limitations for a single twt . . .	5-10
Figure 5-2	Required manufacturing tolerances in twt as a function of frequency	5-11
Figure 5-3	Power characteristics of available low-power backward-wave oscillators. Sperry and Bendix use Karp structure, and CSF uses the vane line. In general, these tubes use low operating voltages and are suited for local oscillator and laboratory source use	5-12
Figure 5-4	Power characteristics of available low-power klystrons. All tubes are reflex klystrons and are light in weight since they use no magnetic focusing fields. They are particularly suited as pump, local oscillator, and laboratory signal sources	5-13
Figure 5-5	Power characteristics of available high-power cw sources. Most significant is 35 percent efficiency achieved in the 813H and marked increase in available power which has occurred since 1960. The letter - A stands for amplifier and the letter - O signifies oscillator	5-16

Figure 5-6	State of the art of the power output of primary transistor oscillators and transistor varactor-multipliers versus output frequency . . .	5-19
Figure 5-7	Typical P-N junction in reverse bias (conditions at avalanche)	5-21
Figure 5-8	Equivalent circuit for avalanche diode at small transit angle	5-21
Figure 5-9	Frequency dependence of the real and imaginary parts of the diode impedance	5-21
Figure 5-10	IMPATT performance	5-23
Figure 5-11	Gunn oscillator performance	5-27
Figure 5-12	Figure of merit for spaceborne cw transmitter .	5-29
Figure 5-13	Figure of merit for ground based cw transmitter	5-30
Figure 5-14	Schematic energy level diagram and processes for singly ionized atoms	5-34
Figure 5-15	Gain curves for right and left hand circular polarization as a function of magnetic field strength	5-36
Figure 5-16	Index of refraction for right and left hand components with and without magnetic field . .	5-38
Figure 5-17	Typical gain saturation characteristic for a laser amplifier	5-40
Figure 5-18	Comparison of (a) inhomogeneous interaction (hole burning) and (b) homogeneous interaction with a single frequency	5-41
Figure 5-19	Nonoverlapping holes burned in the gain line by a multimode laser	5-42
Figure 5-20	Gain line "burned off" by a multimode laser . .	5-42
Figure 5-21	Schematic representation of the change in power with different discharge and cavity lengths	5-45
Figure 5-22	Fractional stability required as a function of desired absolute stability	5-56
Figure 5-23	Fractional stability resulting from a given temperature fluctuation, δT, for several common materials suitable for cavity fabrication	5-60
Figure 5-24	Schematic representation of an optical discriminant	5-63
Figure 5-25	Typical feedback laser frequency control system	5-63

Figure 5-26	Typical Fabry-Perot transmission characteristic. Numerical values typical of gas lasers are given in parentheses	5-65
Figure 5-27	Simple stabilized laser using a Fabry-Perot cavity as a reference	5-67
Figure 5-28	Discriminant characteristics	5-67
Figure 5-29	Dither method of locking a laser to a Fabry-Perot cavity resonance	5-67
Figure 5-30	Discriminant characteristics	5-68
Figure 5-31	Two-beam interferometer balanced discriminator frequency stabilizing system (after Kaminow[3])	5-69
Figure 5-32	Optical discriminant obtained from a two-beam interferometer with path length difference of $L_a - L_b$	5-72
Figure 5-33	Schematic representation of the laser transition gain and power output profiles, showing line center, ν_o, doppler width, $\Delta\nu_D$, and loss line which determines threshold of oscillation	5-72
Figure 5-34	Discriminant derivation	5-73
Figure 5-35	Discriminant characteristics	5-76
Figure 5-36	One variation of the Zeeman effect derived discriminator stabilization system	5-77
Figure 5-37	Energy level diagram showing pertinent levels in CO_2 and N_2	5-80
Figure 5-38	Variation of laser power per unit volume of discharge and laser power per unit length of discharge with discharge diameter	5-89
Figure 5-39	Variation of axial electric field with discharge tube diameter	5-89
Figure 5-40	Variation of optimum discharge current with discharge tube diameter	5-89
Figure 5-41	Schematic illustration of pumping and detection scheme for sun-pumped laser	5-100
Figure 5-42	A cone-sphere condenser	5-100
Figure 5-43	Experimental optical setup for pumping laser rod and detecting laser output	5-100
Figure 5-44	Comparison of lasers for spacecraft transmitter	5-107
Figure 5-45	Comparison of lasers for ground based transmitter	5-107

Figure 6-1	Crystal noise ratio versus local oscillator power	6-5
Figure 6-2	Conversion loss versus local oscillator power and bias	6-5
Figure 6-3	Noise temperature	6-8
Figure 6-4	Detectivity versus wavelength of photoconductive and thermal millimeter and submillimeter detectors	6-15
Figure 6-5	Spectral response of various photocathodes	6-27
Figure 6-6	Signal and noise equivalent circuits of photomultiplier and idealized receiver	6-28
Figure 6-7	Microwave phototube	6-29
Figure 6-8	Physical configuration of electron multiplication system	6-31
Figure 6-9	Three modes of operation of solid state photodetectors	6-34
Figure 6-10	Schematic representation of operating circuits of the photodiode and photoconductor detectors	6-36
Figure 6-11	Representative current-voltage characteristics of a photodiode (upper) and a photoconductor (lower) for various values of incident flux	6-37
Figure 6-12	Constant current equivalent circuits of the photodiode (upper) and photoconductor (lower)	6-38
Figure 6-13	Noise equivalent circuit for the photodiode and photoconductor $\overline{v}R^2\Delta f$ represents excess noise arising in the preamplifier and has been neglected in the text	6-41
Figure 6-14	Monochromatic detectivity as a function of wavelength for several extrinsic photoconductors	6-52
Figure 6-15	Blackbody detectivity as a function of temperature for several extrinsic photoconductors	6-53
Figure 7-1	Multielement optical modulator	7-11
Figure 7-2	Sixteen-element KDP crystal modulator	7-11
Figure 7-3	Two-element parallel-strip modulator	7-13
Figure 7-4	Strip-line traveling-wave modulator configuration	7-15
Figure 7-5	Low frequency one-element optical frequency translator	7-20

Figure 7-6	A possible microwave optical frequency translator	7-20
Figure 7-7	Two-element KDP optical frequency translator for SSBSC modulator . . .	7-21
Figure 7-8	Frequency components at the output of a two-element frequency translator . . .	7-27
Figure 7-9	Cross sectional view of parallel strip transmission line ADP phase modulator .	7-29
Figure 7-10	Modulator for PCM/PL system . . .	7-31
Figure 7-11	Ultrasonic diffraction geometry . . .	7-37
Figure 7-12	Laser oscillator with internal polarization modulation	7-49
Figure 7-13	Laser with coupling modulation by internal frequency shifting	7-50
Figure 7-14	Internal SSBSC modulator	7-51
Figure 8-1	Earth and DSV line-of-sight geometry. DSV-to-Earth LOS shown at instant a signal received from earth. Earth-to-DSV LOS shown as a signal is received from DSV	8-5
Figure 8-2	Deep space vehicle communication control system block diagram for a manned mission	8-10
Figure 8-3	Deep space vehicle communication control system block diagram for an unmanned mission	8-11
Figure 8-4	Unmanned DSV roll control system . .	8-12
Figure 8-5	Unmanned DSV pitch or yaw control system .	8-13
Figure 8-6	Error in predicted pointing angle time between successive versus communications interrogation	8-16
Figure 8-7	Two-axis block diagram of the DSV tracking system	8-18
Figure 8-8	Lead angle, tracking rate, and tracking acceleration curves for a 174 day Mars trajectory	8-19
Figure 8-9	DSV acquisition signal flow diagram . .	8-25
Figure 8-10	DSV transmitter lead angle geometry . .	8-31
Figure 8-11	DSV transmitter lead angle error geometry .	8-31
Figure 8-12	Allowable laser pointing error distribution .	8-32

Figure 8-13	Transmitter lead angle as a function of relative lead angle	8-33
Figure 8-14	Relation of pointing error to track system quantities	8-34
Figure 8-15	Pio's method of pictorial representation of coordinate transformation	8-38
Figure 8-16	Error vector "Pio-gram"	8-39
Figure 8-17	Quadrant detector boresight errors	8-43
Figure 8-18	Assumed weight characteristics of different sized electro-optical apparatus	8-50
Figure 9-1	Bending of an array supported at one fourth the array length from each end	9-4
Figure 9-2	Typical ray in an ideal Luneberg lens	9-11
Figure 9-3	Approximate theoretical electromagnetic field near the focus point of a 10-inch Luneberg lens	9-13
Figure 9-4	Experimental far field radiation pattern of a 10-inch Luneberg lens	9-14
Figure 9-5	Trans-directive array	9-16
Figure 9-6	Specific configuration for a trans-directive array	9-19
Figure 9-7	Van Atta array	9-20
Figure 9-8	Retrodirective array techniques	9-22
Figure 9-9	Electronic beam-forming techniques	9-24
Figure 9-10	Adaptive retrodirective antenna module	9-25
Figure 9-11	Active Van Atta system with separate arrays for reception and retransmission	9-25
Figure 9-12	Modulated Van Atta array	9-26
Figure 9-13	Specific configuration for a retrodirective array	9-28
Figure 9-14	Block diagram of the active adaptive antenna array system simulator	9-28
Figure 9-15	Acquisition time of the 8-channel antenna array simulator versus mixer output voltage	9-29
Figure 9-16	Oscillograms showing the adaptiveness of the phase-locked system	9-29
Figure 9-17	Antenna gain degradation due to random phase errors in aperture illumination	9-32

Figure 9-18	Antenna gain degradation due to random phase errors in aperture illumination . .	9-33
Figure 9-19	Antenna gain degradation due to random phase errors in aperture illumination . .	9-34
Figure 10-1	Diffraction limit for all types of telescopes versus aperture for N_d light	10-2
Figure 10-2	Aberrations of a parabolic reflector off axis, versus f number	10-5
Figure 10-3	Resolution of various types versus field of view	10-6
Figure 10-4	The limiting photo magnitude of a telescope as a function of its aperture and focal ratio	10-9
Figure 10-5	Approximate mirrow weight of beryllium sandwich mirrors	10-10
Figure 10-6	Approximate cost of beryllium sandwich mirrors	10-11
Figure 10-7	Approximate mirror weight of fused silica sandwich mirrors.	10-12
Figure 10-8	Approximate cost of fused silica sandwich mirrors	10-13
Figure 10-9	Combined defocussing and wave front errors	10-14
Figure 10-10	Transmission curve of the Lick 36-inch refractor and its photographic correcting lens	10-17
Figure 10-11	Reflectivity of evaporated aluminum and chemically deposited silver	10-18
Figure 10-12	The reflectances of aluminum and silver for different angles of incidence, when the plane of polarization is parallel (R_p) and perpendicular (R_s) to the plane of incidence .	10-19
Figure 10-13	Preliminary weight estimate	10-20
Figure 10-14	Preliminary cost estimate	10-22
Figure 10-15	Components of a Cassegrain system . .	10-23
Figure 10-16	Tolerance on axial position of the secondary mirror of a Cassegrain telescope . . .	10-24
Figure 10-17	Focal ratio-primary mirror	10-25
Figure 11-1	Power system weight comparison (1 kw system)	11-2
Figure 11-2	Fuel battery system specific energy versus mission duration	11-4

Figure 11-3	Comparison of oriented planer solar cell array and RTG specific weights (power: 25 w to 1 kw).	11-13
Figure 11-4	Battery power vs weight for various discharge times AgZo secondary battery, 25°C temperature	11-14
Figure 11-5	Battery power vs weight for various discharge times AgCd secondary battery, 25°C temperature	11-15
Figure 11-6	Power conditioning subsystem weight vs output power	11-16
Figure 11-7	Relative RTG electrical power output (watts) versus time	11-17
Figure 11-8	Plutonium 238 RTG power vs weight and volume	11-18
Figure 11-9	Typical RTG power vs required radiator fin area	11-19
Figure 11-10	RTG-battery power system unregulated power vs weight	11-21
Figure 11-11	RTG-battery power system unregulated power vs weight	11-22
Figure 11-12	Comparison of nuclear thermionic systems .	11-25
Figure 12-1	Map of the radio sky background at 64 mc (after Hey, Parsons, and Phillips). The contours give the absolute brightness temperature of the radio sky in degrees Kelvin	12-10
Figure 12-2	Map of the radio sky background at 81 mc (after Baldwin). The contours give the absolute temperature of the radio sky in degrees Kelvin.	12-10
Figure 12-3	Map of the radio sky background at 100 mc (after Bolton and Westfold). The contours give the absolute temperature of the radio sky in degrees Kelvin	12-10
Figure 12-4	Map of the radio sky background at 160 mc (after Reber). The contours give the absolute brightness temperature of the radio sky in degrees Kelvin	12-10
Figure 12-5	Map of the radio sky background at 250 mc (after Ko and Kraus). The contours give the absolute brightness temperature of the radio sky in degrees Kelvin	12-11

Figure 12-6	Map of the radio sky background at 480 mc (after Reber). The contours give the absolute brightness temperature of the radio sky in degrees Kelvin	12-11
Figure 12-7	Map of the radio sky background at 600 mc (after Piddington and Trent). The contours give the absolute temperature of the radio sky in degrees Kelvin	12-11
Figure 12-8	Map of the radio sky background at 910 mc (after Denisse, Leroux, and Steinberg). The contours give the absolute brightness temperature of the radio sky in degrees Kelvin.	12-11
Figure 12-9	Galactic noise temperature	12-13
Figure 12-10	Galactic noise along the ecliptic at three frequencies. The 600-mc peak extends to a height of 21 units and the 81-mc peak to 27 units	12-14
Figure 12-11	Radio stars in 378 mc	12-15
Figure 12-12	The most prominent radio stars superimposed on the radio sky background map at 250 mc. (Brightness temperatures of the isophotes are in degrees Kelvin).	12-16
Figure 12-13	Spectra of the most prominent radio stars .	12-17
Figure 12-14	Sun noise temperature versus frequency .	12-20
Figure 12-15	The predicted intensity of planetary thermal radio radiation when the planet is near closest approach to the earth (dashed lines), and the measured intensity of radiation from the planets, the moon, and the nonthermal radio source, Cassiopeia A (solid lines and points).	12-22
Figure 12-16	The brightness temperature of the center of the disk of the moon as a function of the phase of solar illumination. Long dashes—brightness temperature corresponding to the infrared radiation as approximated by the calculated surface temperature of Wesselink (Bull. Astron. Inst. Neth., Vol. 10, pp. 351-363; April 1948). Solid line—brightness temperature corresponding to the 8.6-mm radiation measured by Gibson (proc. IRE, vol. 46, pp. 280-286; January 1958). Short dashes—brightness temperature corresponding to the 3.15 cm radiation measured by Mayer, McCullough, and Sloanaker "Planets and Satellites,"	(cont)

Figure 12-16 (continued)
University of Chicago Press, Chicago, Ill., Ch. 12; 1961). Reproduced from "Planets and Satellites," G. P. Kuiper and B. M. Middlehurst, Eds., University of Chicago Press, Chicago, Ill., 1961; copyright 1961 by the University of Chicago . . 12-24

Figure 12-17 Sky noise temperature due to oxygen and water vapor 12-27

Figure 12-18 Spectral irradiance of brightest stars outside the terrestrial atmosphere . . . 12-27

Figure 12-19 Solar spectral irradiance outside the earth's atmosphere 12-29

Figure 12-20 Calculated planetary special irradiance outside the terrestrial atmosphere . . . 12-32

Figure 12-21 Full moon spectral irradiance outside the terrestrial atmosphere. 12-33

Figure 12-22 Spectral albedo versus wavelength . . . 12-33

Figure 12-23 Position of Mercury and Venus at greatest elongation ($\alpha = 90^{o}$) 12-34

Figure 12-24 Solar and terrestrial radiation. Reflected solar and total earth radiation to space values should be divided by π to obtain the radiance for each case. T_S is the surface temperature and T_A is the effective radiating air temperature 12-36

Figure 12-25 Illumination factor as a function of altitude for parametric values of bistatic angle, σ . 12-38

Figure 12-26 Clear sky radiance at Colorado Springs, Colorado; sun angle = 45 degrees . . . 12-39

Figure 12-27 Diffuse component of typical background radiance from 30 km and sea level zenith angle 45 degrees and excellent visibility 12-40

Figure 12-28 Estimated average diffuse radiation from zenith night sky. Average diffuse radiation from zenith, 0.3 to 1.3 microns . . 12-42

Figure 12-29 Spectral radiance in the infrared of a clear sky for several elevation angles above the horizon. These spectra were measured at an elevation of 11,750 feet, at night, with an ambient temperature of $8^{o}C$ 12-43

Figure 12-30	Zenith twilight intensity	12-44
Figure 12-31	Zenith moonlight intensity	12-45
Figure 12-32	Correction factor for moon phase	12-46
Figure 12-33	Ionospheric attenuation for a source at 1000-km height	12-48
Figure 12-34	One-way attenuation through the standard summer atmosphere due to oxygen and water vapor	12-48
Figure 12-35	Total attenuation for one-way transmission through the atmosphere. (Proc. IEEE, p. 488, April 1966)	12-49
Figure 12-36	Atmospheric attenuation summary	12-50
Figure 12-37	Atmospheric distortion of RF wavefront based on NBS Maui data	12-52
Figure 12-38	Atmospheric transmission, 0.3 to 1.3 microns	12-52
Figure 12-39	Atmospheric transmission, 1.2 to 5.0 microns	12-53
Figure 12-40	Fog-rain attenuation, $\lambda = 0.63\mu$	12-56
Figure 12-41	Power reduction caused in an optical heterodyne receiver caused by loss of phase coherence (after Fried)	12-56
Figure 12-42	Dependence of coherence length, γ_o on zenith angle, ϕ, and wavelength λ_μ (after Fried)	12-58
Figure 12-43	Geomctry for optical heterodyne receiver in the atmosphere (after Fried)	12-58
Figure 12-44	Dependence of the amount of quivering of stellar images on the zenith distance	12-60
Figure 12-45	Empirical dependence of the amount of star twinkling on the diameter of the telescope diaphragm (1, winter; 2, summer)	12-62
Figure 12-46	Dependence of the amount of star twinkling on the zenith distance when the telescope diaphragm has a diameter of three inches	12-62
Figure 12-47	Dependence of the amount of star twinkling on the zenith distance when the telescope diaphragm has a diameter of 12.5 inches	12-62
Figure 12-48	Mean-square deviation in log-level of intensity, X_I^2	12-63

Figure 12-49 Mean-square deviation on log-level of intensity, X_I^2 12-63

Figure 12-50 Mean-square deviation in log-level of intensity, X_I^2 12-65

Figure 12-51 Geometry of propagation through a turbulent blob 12-65

Figure 12-52 Mean-square variation (Δn^2) versus altitude, h 12-67

Figure 12-53 Correlation distance, L_o versus altitude h 12-67

Figure 13-1 Limiting visual magnitude versus aperture . 13-2

Figure 13-2 The relative sensitivity at various wavelengths of (A) the light-adapted eye and (B) the dark-adapted eye 13-3

Figure 13-3 MSFN and STADAN visibility contours for typical orbital altitude of 200 nm at 5 degrees elevation 13-26

Figure 14-1 Correlation of Lockhart and Martinelli for two-phase flow 14-4

Figure 14-2 Radiative heat flux versus temperature . . 14-8

Figure 14-3 Effective radiative heat flux versus temperature of radiators in direct sunlight at various solar distances . . 14-9

Figure 14-4 Fin effectiveness versus radiation modules . 14-10

Figure 14-5 Radiator area requirements for subcooler section 14-12

Figure 14-6 Emissivity versus temperature (Data from NACA-TN-4206) 14-13

Figure 14-7 Specific weights of alkali metal direct condensing radiators 14-14

TABLES

Table 3-1	Apollo applications manned earth-orbit flight program.	3-11
Table 3-2	Communications subsystem summary table	3-12
Table 3-3	Planetary fly-by scientific payload	3-14
Table 3-4	Typical data requirements	3-15
Table 3-5	Voyager spacecraft preliminary designs.	3-18
Table 3-6	Typical mission durations	3-25
Table 4-1	Detection techniques	4-2
Table 4-2	Laser modulation techniques	4-19
Table 4-3	Source bit rate reduction	4-33
Table 5-1	Negative grid tube characteristics	5-3
Table 5-2	Type classification of microwave tubes - oscillators and amplifiers	5-4
Table 5-3	Klystron characteristics	5-5
Table 5-4	Traveling wave tube characteristics	5-6
Table 5-5	Crossed-Field tube characteristics	5-7
Table 5-6	Spontaneous radiative transition probabilities in CO_2	5-81
Table 5-7	Continuous wave laser oscillation wavelengths in the $00^0 1$ to $10^0 0$ band of CO_2	5-87
Table 5-8	State-of-the-art CO_2 lasers (May 1966)	5-91
Table 5-9	CW laser oscillators	5-92
Table 5-10	Typical high energy laser oscillators	5-96
Table 5-11	High peak power laser oscillators	5-97
Table 5-12	High repetition frequency laser oscillators	5-98
Table 5-13	Figure of merit for several gas lasers	5-106
Table 6-1	Detector characteristics	6-16
Table 6-2	Quantum efficiencies and responsivities of red-sensitive photosurfaces	6-46
Table 6-3	Wavelengths of peak response and Engstrom's factors	6-48
Table 6-4	Noise in photomultiplier tubes	6-48
Table 6-5	NEP in one-cycle band at 6328 A and 6943 A for several photomultipliers	6-49

Table 6-6	Dynamic resistance and detectivities of candidate photodiode materials	6-50
Table 6-7	Characteristics of semiconductor materials for 10.6μ detector	6-51
Table 7-1	Optical and electrical constants of KDP	7-8
Table 7-2	Properties of some typical electro-optic materials.	7-9
Table 7-3	Elasto-optic performance parameters for selected crystals.	7-46
Table 7-4	Characteristics of some electro-optic modulators - June 1966	7-54
Table 8-1	Stabilization and control system torquing techniques summary	8-20
Table 8-2	Attitude sensing techniques summary	8-22
Table 8-3	Star trackers.	8-47
Table 10-1	Effect of wave front errors in energy density at the center of the beam.	10-3
Table 10-2	Optical materials	10-18
Table 11-1	Fuel cell battery specifications.	11-3
Table 11-2	Solar Cell System with Regenerative Fuel Cell Energy Storage, Pounds (System weight)	11-6
Table 11-3	Primary space batteries	11-8
Table 11-4	Solar array degradation in Earth-Mars transit by solar flare activity n/p silicon cells with 30 mil quartz covers	11-11
Table 11-5	Solar array degradation in Earth-Mars transit by solar flare activity n/p silicon cells with 45 mil quartz covers	11-12
Table 11-6	Typical radiation limitations for unmanned spacecraft	11-20
Table 11-7	Performance of reactor thermionic power system design with converters in heat exchanger (1 megawatt electrical output)	11-27
Table 11-8	Performance of reactor thermionic power system design with converters in radiator (1 megawatt electrical output)	11-28
Table 11-9	Effect of thermionic converter performance on system performance for design with integral thermionic fuel element (1 megawatt electrical output).	11-29

Table 11-10	Comparison of several nuclear thermionic concepts for generation of space vehicle electrical power (1 megawatt electrical output)	11-30
Table 12-1	Planetary radiation	12-21
Table 13-1	Scale for rating seeing	13-2
Table 13-2	Baker Nunn sites showing percent of time lost due to clouds	13-6
Table 13-3	Astronomical telescopes of 36 inches or greater aperture	13-7
Table 13-4	Residual spherical aberration of large telescopes	13-8
Table 13-5	Station location	13-9
Table 13-6	210-ft altazimuth antenna: expected performance	13-11
Table 13-7	DSIF frequencies	13-13
Table 13-8	DSIF receiver capabilities	13-15
Table 13-9	Network evolution.	13-18
Table 13-10	Apollo MSFN schedule.	13-19
Table 13-11	LEM down-link MSFN S-band transmission summary	13-22
Table 13-12	CSM to MSFN S-band transmission combination summary (FM modes)	13-23
Table 13-13	CSM to MSFN S-band transmission combination summary (PM mode)	13-24
Table 13-14	LEM up-link S-band transmission combinations summary	13-25
Table 13-15	MSFN to CSM S-band transmission combinations summary	13-25
Table 13-16	STADAN ground stations system equipment	13-27

1.0 INTRODUCTION

1.1 PURPOSE

"Reference Data for Advanced Space Communications and Tracking Systems" is being produced for the Goddard Space Flight Center by the Hughes Aircraft Company under NASA Contract NAS 5-9637. The purpose of this contract is to perform a parametric study with the following goals:

1. Perform overall systems trade-off studies in sufficient detail to identify those missions which will made the best use of laser/optical, microwave, or a combination of microwave and laser/optical communication and tracking systems.
2. Provide a plan for optimumly integrating such future microwave and/or laser/optical communication and tracking systems into present and future world-wide systems.
3. Provide overall systems design criteria or specifications for microwave and/or laser/optical communication and tracking systems.

The first issue of "Reference Data for Advanced Space Communication and Tracking" provided preliminary data for the achievement of these goals. This data was presented in thirteen technical sections, in addition to the introductory material. These technical sections, were designed to meet the contract goals, listed above, as follows:

- The methodology section provides a basis for determining the optimum communication and tracking system configuration, whether it be at microwave, millimeter waves, or optical frequencies. The methodology section, in its detailed documentation of functional relationships, also establishes the parametric studies which must be performed in each technological area.

- Two sections of the Reference Data will dominate in establishing a means of integrating future space communication systems with existing and future ground facilities. These are Mission Analysis and Ground Receiving Sites. Mission Analysis is concerned with overall future space mission goals and the facilities for their achievement. Present Ground Receiving Sites are documented in the first issue. Subsequent issues will refine this documentation and establish criteria for future receiving systems.
- The basis for system design criteria are found in the remaining technical sections. They are: Communication Theory, Transmitting Power Sources, Detectors, Optical Modulators, Acquisition and Tracking, Radio Frequency Antennas, Optics, Spacecraft Prime Power Generation, Background Radiation and Atmospheric Attenuation, and Heat Transfer Systems.

This is the second issue of "Reference Data for Advanced Space Communication and Tracking System." In this issue five of the technical sections of the first or preliminary issues are updated and extensively expanded.

The sections given in this issue completely replace their corresponding sections in the first issue. The sections which are updated and enlarged in this second issue are: Mission Analysis, Transmitting Power Sources, Optical Modulators, Acquisition and Tracking, and Background Radiation and Atmospheric Attenuation. Updating and enlargement of other sections will follow the schedule given in the Summary section.

1.2 REFERENCE DATA FORMAT

Each Reference Data section is formed of six basic subsections. They are:

- Introduction
- Theory
- Performance
- Burden Relationships
- Nomenclature
- References

Each subsection of this basic structure is used in the component technology sections. However, some sections are not completely amenable to this organization of material and variations are allowed. The intent of the subsections is as follows.

1.2.1 Introduction

This subsection introduces the material of the section. The status of the section in relation to its ultimate development may be noted.

1.2.2 Theory

This subsection is designed to introduce the reader to the theory of the technological area being discussed. Basic relationships are given but extensive derivations are avoided. The theory is presented as a guide for using the material of the section and as a means to project future parameter capabilities.

1.2.3 Performance

This subsection contains the documented state of the art of the technology. It lists new varients of the technology and tabulates parameters and performance.

1.2.4 Burden Relationships

This subsection contains the parametric relationships of the section technology. Relationship of particular concern is parameter

values as a function of weight, cost, size, etc. Ancillary equipment required by the technology are also described in this subsection.

1.2.5 Nomenclature

A nomenclature subsection has been introduced with this issue of "Reference Data for Advanced Space Communication and Tracking Systems." It is intended to apply only to the section of which it is a part. It will form the basis for a complete and uniform nomenclature for the entire volume. This should be realized in the 6 December 1966 issue.

1.2.6 References

This subsection lists references used in the section proper. The references are not intended as an extensive bibliography but rather to direct the reader to the source of the documented material.

1.3 FRONT MATTER

The Table of Contents, List of Illustrations and List of Tables is complete for the entire volume of "Reference Data for Advanced Space Communication and Tracking Systems." The date of last issue for each section is given in the Table of Contents. Text material for sections not updated in this issue may be found in the first issue dated 6 February 1966, Report Number P66-16.

3.0 MISSIONS ANALYSIS

3.1 INTRODUCTION

The choice of a space communication system for a particular mission must take into account

1. Probable objectives of the mission under consideration
2. Reflection of these mission objectives into communication system requirements.

This involves definition of communication range and system lifetime requirements and their effects on data transmission rate and data processing and storage facility requirements. More basic limitations are the restrictions on communication system weight, volume, power, aperture size and pointing accuracy imposed by technological limitations.

3.2 PROBABLE MISSION OBJECTIVES

Earth orbital, lunar, and planetary missions envisioned over the next several decades and their approximate inception dates are summarized in Figures 3-1 through 3-4.[1] Objectives of these and other missions were described by NASA Administrator Webb in his letter[2] to President Johnson of May 1964 under the following headings:

3.2.1 Earth Orbital Operations

A. Unmanned Scientific Missions

1. Investigations of the Sun, the space environment, and their interrelationship.

a. Orbiting solar observatories carrying telescopes and other instruments above the atmosphere to study the electromagnetic radiation from the sun.

b. Earth satellites to study specific phenomena such as the influence of the sun on the earth's atmosphere, the particle trapping mechanism and the interaction of the solar wind and the earth's magnetic field.

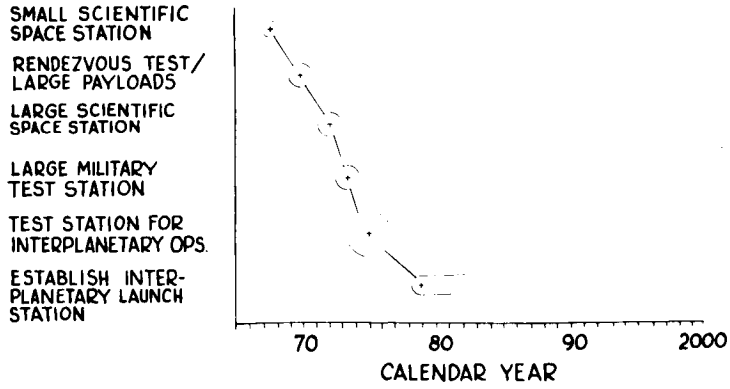


Figure 3-1. Orbital missions.

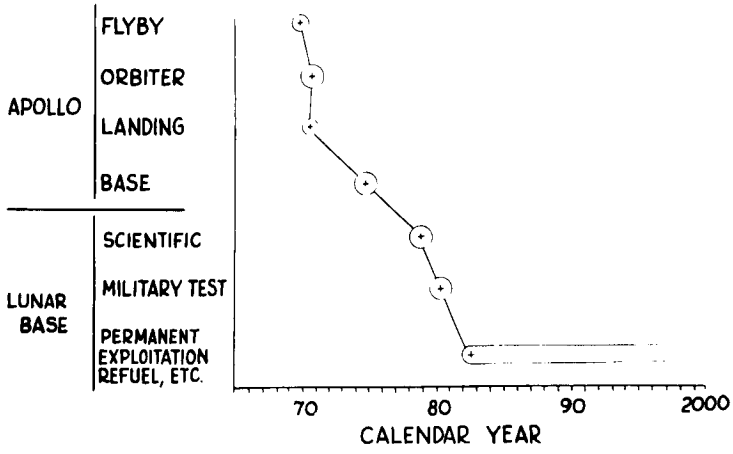


Figure 3-2. Lunar missions.

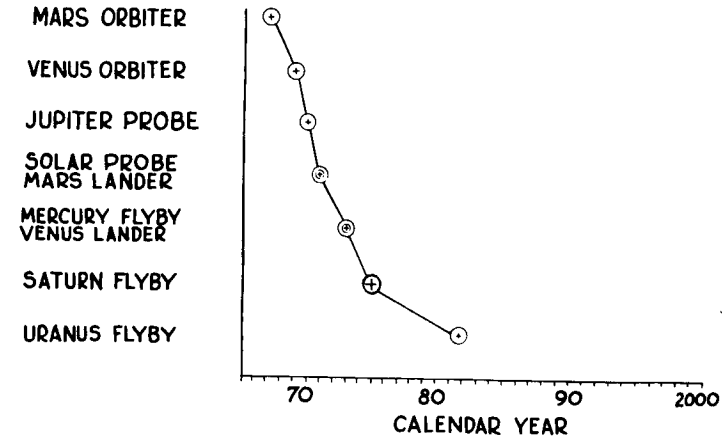


Figure 3-3. Unmanned planetary missions.

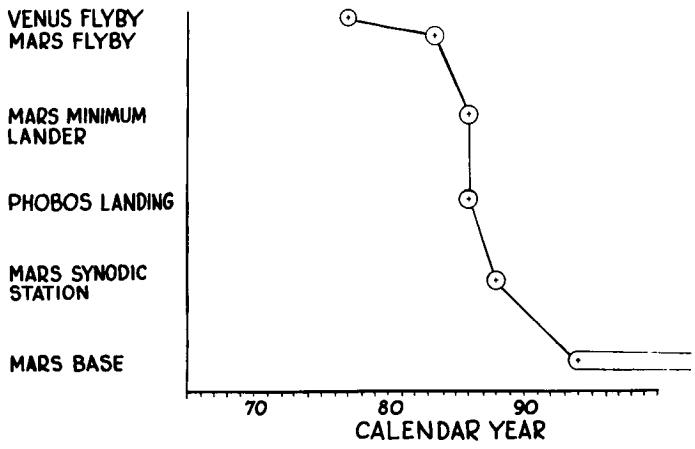


Figure 3-4. Manned planetary missions.

c. Interplanetary probes to study the particulate radiation from the sun, to determine the solar atmosphere, and to investigate the space environment.

2. Astronomical investigations - Orbiting astronomical observations to study the electromagnetic radiation emitted by the stars and galaxies from above the absorbing and disturbing influence of the atmosphere.

3. Biological investigations - Recoverable satellites carrying animal and plant life to study the effect of the space environment, or the absence of terrestrial effects, on living organisms; and specially instrumented probes to search for evidence of extraterrestrial life forms in meteorites and asteroids.

B. Manned Operations

1. Spacecraft to orbit three men for periods up to 2 weeks to further extend knowledge of man's capabilities in space and to develop the rendezvous, docking, and other flight operations necessary to carry out the manned lunar mission. (Apollo).

2. Spacecraft to orbit two men for a period of 30 days to investigate the potential of earth-orbital flight for future military operations. (Manned Orbiting Laboratory).

3. Spacecraft to orbit two or three men for periods up to 90 days to obtain fundamental knowledge on the performance of man in longer duration space missions to accomplish manned scientific and engineering research tasks, and to develop the spacecraft capability for longer duration lunar missions and for logistics resupply of possible earth orbital laboratories. (Apollo Applications).

4. Orbital manned laboratories

a. Laboratories capable of sustaining nine to twelve men and of operating for periods up to 1 to 3 years with appropriate logistics resupply in order to utilize man's abilities to observe, to reason, and to perform complex tasks, such as: to conduct a wide variety of research activities, to develop operational capabilities in earth orbit, to develop and qualify subsystems and systems applicable to lunar and planetary missions, and to establish the feasibility of

manned missions of sufficient duration as would be required to accomplish planetary exploration. (Manned Orbiting Research Laboratory).

b. Large laboratories or semi-permanent space stations, capable of sustaining up to 24 men, as major installation in space to accomplish to a greater degree the objectives of the smaller laboratories, with potential for serving as supporting facilities for planetary missions.

c. Manned laboratories in synchronous orbit for the primary purpose of enabling continuous manned observation of major portions of earth.

C. Applications of Space Flight Technology

1. Meteorological Satellites

a. Extension of the Nimbus capabilities incorporating improved sensors, camera, control, and data systems to permit remote measurement and readout of atmospheric data at several altitudes over the entire globe several times a day.

b. Spacecraft orbiting at synchronous altitude to develop a capability for continuously viewing fixed areas of the earth in order to continuously monitor transient weather phenomena such as thunderstorms, fronts, or squall lines.

2. Communications and Navigation Satellites - Advanced satellites to develop that technology needed to transmit from satellites directly to small land and airborne receiving sites equipped with simple, economical antennas. Development of this technology could lead to systems for the control and coordination of air traffic in congested trans-ocean air lanes, for all-weather global navigation and position fixing; and for direct radio and television broadcasting from satellites to home receivers.

3.2.2 Lunar Exploration

1. Manned exploration of the lunar surface for periods of approximately 2 weeks or more utilizing a logistics system to provide

the necessary shelter, life support, and roving vehicles to enable scientific studies of selected areas within a range of several miles from the landing point.

2. Establishment of semi-permanent lunar stations to conduct extensive scientific investigation of larger areas on the moon through the development and utilization of larger logistics systems.

3. Establishment of permanent lunar bases for the exploitation of lunar resources, for staging of deep space missions, or as major scientific observation sites.

3.2.3 Planetary Exploration

Specific planetary missions that are currently planned and those that are under consideration are as follows:

A. Unmanned Mars-Venus Exploration

1. Spacecraft which fly by the planet with instruments to determine temperatures of the surface and the atmosphere; to obtain information about the atmosphere to aid in the design of vehicles to enter the atmosphere and land; to look for evidence of extra-terrestrial life; to look for magnetic fields, ionosphere and radiation belts.

2. Spacecraft which fly by the planet and send an entry vehicle into the atmosphere to more accurately determine the characteristics of the atmosphere.

3. Spacecraft to orbit the planet and land a capsule on the surface to begin the systematic study of the properties of the atmosphere, ionosphere, and magnetosphere of the planet and the search for extraterrestrial life.

4. Spacecraft to land an automated biological laboratory to study the biological systems on the planet, to study the effect of the planetary environment on terrestrial biology and to provide significant information needed to design spacecraft for a manned landing on the planet.

The first of these is the successful Mariner program which has now been extended to include a 1967 Venus mission and a 1969 Mars mission.[3] The last three categories of missions are grouped

under the generic name, Voyager, scheduled for 1973 at the earliest. As presently conceived, Voyager will take the form of a combined orbiter with a landing capsule. Orbiter instrumentation will include equipment for continuous measurements of atmospheric phenomena, temperature, and composition. Lander instrumentation would include life detection experiments, surface photography, seismometer and gravity measurements.

B. Manned Mars-Venus Exploration

1. Manned landing missions to Mars in order to obtain scientific knowledge of the atmospheric, surface, and subsurface characteristics of the planet.

2. Manned fly by missions of Mars and Venus to obtain additional fundamental knowledge of the interplanetary environment, the planetary atmospheres, and the planetary surfaces through the use of manned observations and data gathering, as well as unmanned probes and landers carried on board the manned spacecraft.

3. Manned orbiting missions of Venus, and, if possible, land on the surface for the purpose of obtaining scientific knowledge of the planet.

C. Unmanned exploration of the distant planets, asteroids and comets

1. Spacecraft which fly by Jupiter to study the magnetic fields and trapped radiation belts believed to be responsible for the radio noise from Jupiter.

2. Spacecraft to orbit Jupiter to map the magnetosphere, determine structure, composition and temperature of the atmosphere.

3. Spacecraft to fly by Mercury to determine the surface characteristics and measure the surface temperature.

4. Spacecraft to fly by asteroids to ascertain the nature and surface characteristics of the asteroids and to attempt to determine whether they are the remains of a planet or have another origin.

5. Spacecraft to fly through the tail of a comet to determine the composition of a comet, its origin and the physical laws which govern cometary behavior.

The major forthcoming new interplanetary mission is Voyager. NASA and JPL are presently working on in-house conceptual design studies while Boeing, TRW, and GE have completed studies of Voyager. It is expected that a preferred design approach will be arrived at by November 1967 and a single contractor selected by April 1968. Present plans call for use of a Saturn V booster and spacecraft weight in the 10,000- to 20,000-pound range. The first Voyager launch has been delayed until at least 1973.[3]

Manned missions beyond Apollo are summarized in Figure 3-5. NASA presently envisions a 4-year program of Apollo Application missions listed in Table 3-1. These will probably be followed by the Manned Orbiting Research Laboratory with a crew of 9 to 12 men. Apollo Applications missions for lunar exploration include both orbiters and landers. The first manned planetary mission sometime in the 1975-1985 period will certainly be a fly-by, rather than a lander mission, with Mars as the most probable objective.[4]

3.3 COMMUNICATION SYSTEM REQUIREMENTS FOR PROBABLE MISSIONS

Typical communication system problems and limitations for various missions are listed in Table 3-2 (reproduced from the NASA Pattern Relevance Guide).[5]

3.3.1 Data Transmission Requirements

Data transmission requirements depend on the number and type of experiments to be carried, the time interval during which information must be returned, and the available information storage capacity. Typical instrument payloads for an unmanned planetary fly-by spacecraft and their associated data outputs are listed in Table 3-3.[6] The optimum transmission rate for returning this data must be determined by a trade-off between transmitter power, transmitting and receiving aperture size, information storage capacity and reliability considerations. Typical maximum data rate requirements for various types of information are listed in Table 3-4. Real-time television data rate

	1965	1966	19	1968	1969	1970
MAN-RATED BOOSTERS	MAN-RATED TITAN 2		MAN-RATED TITAN 3C			
		SATURN 1B				
			SATURN 5			
MANNED MISSIONS	GEMINI	APOLLO (EARTH ORBIT)		MOL	APOLLO (LUNAR LANDING)	ORBIT LOGISTICS
				AES MANNED EARTH ORBIT (PRIMARY FLIGHT MISSIONS)		
				• BIOSCIENCE EXPTS • EXTRAVEHICULAR ENGINEERING ACTIVITIES (EVEA) • PHYSICAL SCI BIOMED/BEHAV	• TEST LUNAR ORBIT SENSORS • ATMOS SCI EXPTS; ORBITAL TECH DEV • ADV TECH R&D	• BIOMED/BEHAV EXPTS • BIOMED/BEHAV EXPTS; EARTH SCI & RESOURCES • PERFECT CREWING AND OPNS (INITIAL 45-DAY MISSION)
						AES MANNED LUNAR ORBIT SUR
						• LUNAR SURFACE RECON/MAPPING
						AES LUNAR LANDING MISSION
						• INVESTIGATE SURFACE FEATURES (1-MAN TRANSPORT • FOLLOW-UP MISSION (EXTENSIVE SURFACE TRAVERSE)
ADVANCED SYSTEMS			EARTH ORBITAL OPERATIONS (RENDEZVOUS, LOGISTICS)	REUSABLE PAYLOADS	AUTONOMOUS NAVIGATION	UTILITY KICK STAGE
					LEM TRUCK	
PACING TECHNOLOGIES	EXTENDED DURATION ECOLOGICAL COMPONENTS (OPEN/CLOSED CYCLE)	"1-YEAR" ELECTRONICS	AUXILIARY PAYLOAD TECHNIQUES AUTOMATIC CHECKOUT FUEL CELLS (HOURS) RENDEZVOUS AND DOCKING: RADAR, COUPLING MECHANISMS LONG-TERM ATTITUDE CONTROL LUNAR ENVIRONMENT PROTECTION	ZERO-G PROPELLANT CONTROL AND MEASUREMENTS "SPACE" METALLURGY RUGGED PRECISION OPTICS METEOROID PROTECTION STABILITY OF LARGE STRUCTURES IN LOW-G ORBITAL LASERS, MASERS REMOTE SENSORS PHOTOGRAPHIC DATA PROCESSING LUNAR SURFACE POWER SUPPLY (MONTHS)	SPACE-STOREABLE PROPELLANTS EXPANDED LAUNCH OPERATIONS LUNAR SURFACE LOCOMOTION (ONE MAN)	MATERIALS RESISTANT TO LONG-TERM SPACE "SOAK" NEW MATERIALS, APPLICATIONS ADVANCED FLIGHT MECHANICS LONG-TERM EVA WORKER FACILITIES

3-9-1

1971	1972	1973	1974	1975	1976-1980
		POST-AES ORBITAL SPACE STATIONS (MORL-TYPE, 1-5 YEAR, 10-20 MAN)			NATIONAL MULTI-PURPOSE SPACE STATIONS (NMSS, 3-5 YEAR, 12-36 MAN) EARLY PLANETARY MISSIONS MARS FLYBY VENUS FLYBY
ASTRONOMY/ASTRO-PHYSICS • SYNOPTIC WEATHER SURVEY • EVEA (PROP. HANDLING, RECAPTURE SYNCOM 3)	•EARTH SCI & RESOURCES •ASTRONOMY/ASTRO-PHYSICS • PRECURSOR GLOBAL COMM •ARTIFICIAL G/COMPL BIOMED EXPTS				
/EY					
• POLAR ORBIT RECON	• SURFACE MISSION SUPPORT LAB				
IS					
• 14-DAY EXPEDITION (ADV SURFACE TRANSPORT)	•ESTAB SEMI-PERM ASTRO-PHYSICS/ASTRO LAB •VERIFY EARLIER GEOPHYSICS DATA: •TEST ADV SURFACE SHELTER	•COMPLETE SCI AND TECHNOLO-GICAL EXPTS			6-MONTH LUNAR EXPEDITION PERMANENT LUNAR BASE
MOLAB	REUSEABLE STAGES	EARTH ORBIT OPERATIONS (BOOSTER BUILDUP)		NUCLEAR SOLID-CORE SYSTEMS	CLOSED ECOLOGICAL SYSTEMS NUCLEAR ELECTRIC SYSTEMS RECOVERABLE BOOSTERS
	SATURN-NUCLEAR INTEGRATION ON-BOARD SELF CHECKING ELEC,-TRONICS CREW AND EQUIPMENT SHIELDING LONG-DISTANCE COMMUNICATIONS (MARS/VENUS) ORBITAL LAUNCH TECHNIQUES	COMBINED ANTI-METEOROID/-INSUL-ATION STRUCTURES LONG-LIFE ELECTRONICS (10-15 YEAR) SPENT-STAGE UTILIZA-TION IN ORBIT INTERPLANETARY NAVIGATION TECH-NIQUES			PROPELLANT AND GAS TRANSFER IN SPACE MARS ENTRY AERO/ RETRO BRAKING LONG-TERM POWER SUPPLIES (YEAR) SHIELDING DATA TRANSMISSION AND HANDLING HIGH-ENERGY SPACE- STORABLE PROPELLANTS

Figure 3-5. Timetable of future manned space missions.

Table 3-1. Apollo applications manned earth-orbit flight program.*

Booster/ Approx. Date	Altitude, Inclination, Duration	Primary Flight Objectives	Secondary Flight Objectives
Saturn 1B Mar. '68	200 nm 28.5 deg 14 days	Bioscience experiments: physiological effects of zero-g on crew and test animals, monitoring of microorganisms in near-earth space.	Perfect orbital operations techniques, advanced mission spacecraft subsystems.
Saturn 1B Sept. '68	200 nm 28.5 deg 14 days	Perfect extravehicular engineering activities (EVEA); emphasis on astronaut locomotion.	Carry basic biomedical/behavioral package (BBP). Environment physics: radiation investigations, Mars spectrometry, test optical materials.
Saturn 1B Nov. '68	200 nm 28.5 deg 14 days	Physical sciences: measure radiation; collect micrometeoroids; study Atlantic geomagnetic anomaly, capillarity, zero-g fluid dynamics, etc.	EVEA assembly techniques, external maintenance and repair, personnel and cargo transfer. Carry BBP; study zero-g physiological effects.
Saturn 5 Apr. '69	200 nm 90 deg 14 days	Validate advanced sensors for earth and lunar-orbit surveillance of atmosphere composition and dynamics; synoptic mapping; multispectral and multifrequency radar terrain analysis.	Study personal hygiene techniques and food technology; experiment with artificial-g devices; develop techniques and equipment for space rescue; perform close-hand recon of satellite (Echo).
Saturn 5 May '69	Sync. 0 deg 14 days	Survey upper atmosphere via i-f radiometry and spectrometry; study extraterrestrial effects on weather (conjugate aurora and airglow).	Evaluate advanced space suits; experiment with small maneuverable satellite; test operation of solar sailing passive cc msat; biomedical/ behavior studies.
Saturn 1B Dec. '69	200 nm 28.5 deg 14 days	Test optical instruments and materials; study advanced LSS and navigation and guidance; EVA manipulation and locomotion.	Operate 38.5-in. reflecting telescope, manned coronagraph, x-ray astronomy; collect micrometeoroids; biomedical/behavior studies.
Saturn 1B Jan. '70	200 nm 28.5 deg 14 days	Perform extensive biomedical and psychological studies in zero-g; validate countermeasures against circulation reflex changes, hemostasis; study sensory perception and complex task performance.	Extensive bioscience studies of limb regeneration and wound healing, otolith function in frogs, effects of space on animal tissue; collect microorganisms in near-space; practice orbit maneuvers and docking.
Saturn 5 June '70	200 nm 83 (retro) 45 days	Extensive biomedical/behavior experiments: determine crew work capacity and metabolic work loads.	Atmospheric studies; radiometry and spectrometry; test prototype startracker; agriculture, forestry, geology, and oceanographic surveillance.
Saturn 1B Oct. '70	200 nm 28.5 deg 45 days	Perfect long-term hygiene and food technology; develop advanced waste management techniques; flight-test small maneuverable satellite.	Study long-term genetic effects on microorganisms and animal tissue; effects of drugs on behavior, on task performance, and on crew work/rest routines; artificial-g.
Saturn 5 Apr.'71	Sync. 0 deg 45 days	Astronomy/astrophysics mission: search for extraterrestrial life, study planetary surfaces and internal mechanisms of astronomical bodies.	Optimize communication relay systems; deploy r-f reflectors, gravity-gradient booms and satellite relay; possibly recapture Syncom 3.
Saturn 5 Aug. '71	200 nm 28.5 deg 45 days	Atmospheric, earth science, resources surveillance mission; multispectral, synoptic earth mapping and radar imaging.	Extensive biomedical/behavior experiments; study space effects on structures; EVA deployment of large solar cell arrays.
Saturn 1B Oct. '71	200 nm 28.5 deg 45 days	Study EVEA procedures for transfer, launch and recovery; develop propellant handling techniques.	Monitor radiation; fundamental capillarity processes; test advanced G&N units; biomedical studies.
Saturn 5 Apr. '72	200 nm 90 deg 45 days	Radiometric and spectrometric studies of upper atmosphere dynamics; test prototype star-tracker; mapping, photo- and radar-imaging of global agriculture, geology and ocean features.	Further experiments with small orbit-launched satellite; perfect advanced orbit maneuvering and docking; study orbit assembly and other logistics operations; flight-test advanced spacecraft subsystems.
Saturn 5 Aug. '72	200 nm 28.5 deg 45 days	Astronomy/astrophysics mission: study planets with 38.5-in. optics; radio-astronomy; operate as manned solar observatory.	Develop technology for advanced LSS; test R&D models of on-board guidance and navigation systems.
Saturn 1B Oct. '72	200 nm 28.5 deg 45 days	Investigate advanced communications/navigation techniques; erect large-aperture antenna.	Various space physics studies: radiation environment, micrometeoroids, airglow horizon photography. Study fluid behavior, pool boiling and nucleate condensation.
Saturn 1B Nov. '72	200 nm 28.5 deg 45 days	Complete biomedical/behavior studies; verify long-term space medicine standards and protective measures.	Evaluate advanced space suits, worker facilities, related EVA equipment; perfect on-board centrifugation.

As presently conceived by NASA, the AES program extends over a period of four years. (For a more detailed breakdown of the experiments for AES flights, see "Status Summary, Manned Experiment Program for AES," NASA-Hq., OMSF, Wash., D.C. Aug. 2, '65.)

Table 3-2. Communications subsystem summary table.

	First Terminal	Second Terminal	Problems and System Limitations
(1)	Earth	Earth	VLF & LF links require high power, depend primarily upon ground wave, large antennas, small bandwidths. MF & VHF links require judicious choice of times when ionospheric reflections permit sky wave paths. UHF, SHF & EHF or optical are limited to line of sight unless scatter can be made "visible" to the receiver. Medium to high bandwidths, small antennas, some atmospheric absorption. Tower-to-tower links are excellent. Cable links are expensive and require several years' lead time although many leasable links exist. Bandwidth usually limited to coaxial cable's high frequency cut-off. Relay satellites are permitting long-distance, wide-band links and such systems are expected to improve rapidly.
(2)	Earth	Orbiting Satellite of the Earth	VLF & LF generally unusable due to ionospheric shielding. MF & VHF occasionally usable when tropospheric ducting is applicable. UHF, SHF & EHF are usable, but due to line-of-sight limitations, satellite must be above horizon. Tracking directional antennas are used to improve receiver and transmitter "gain". Optical links are limited by cloud cover, atmospheric refraction, scattering, and reflection. Doppler shifts must be compensated.
(3)	Orbiting Satellite of the Earth	Earth Receiver	VHF & UHF usually used. Satellite usually has omnidirectional antenna, but may be oriented to local vertical (by attitude control) and use a hemispherical pattern antenna or restricted, conical major lobe. Satellite transmitter is usually power limited, which requires high gain antennas at Earth terminal. (Az-El tracking antennas give best results; Satellite must be above horizon.)
(4)	Earth	Space Probe	Conditions similar to (3) but power and directivity at Earth terminal must be adequate for greatest range to be used. This usually requires extremely large antennas with precision tracking in right ascension and declination at Earth locations. Probe must be above horizon of tracking station, which is a function of elevation angle and attitude of the probe. Three DSIF stations can handle probes in the ecliptic plane - one at a time - but many probes aloft at once may require either undesirable time sharing or construction of new DSIF facilities.
(5)	Earth	Vehicle on Moon or Moon Base (on near side)	Conditions similar to (4) above. Possibility exists for deploying large antennas on lunar surface or vacuum plating crater walls to provide semi-directional "dishes" facing toward Earth. Eventual installation of extensive antenna "farms" on the moon may permit use of lower frequencies for direct broadcast to Earth receivers (such as "TV from the moon").
(6)	Moon Base #1 (near side) or Lunar Roving Vehicle	Moon Base #2 (near side)	Probable lack of lunar ionosphere requires either high power VLF ground wave system, UHF line-of-sight (tower-to-tower for long distance), or moon-Earth-to-moon relay or use of a lunar satellite relay (VHF & UHF or optical). Cables may be strung (eventually) between fixed bases. Electron guns may be used to "orbit" a curved beam of modulated current at low altitude and directed along a great circle between bases to achieve circumlunar communications, such as between observatory on far side and relay station on near side (feasibility not yet proved). A high "launch" tower is required for the "gun".
(7)	Moon Base	Orbiting Satellite (of Earth's Moon)	Satellite must be over horizon. Directional tracking antennas preferred unless nuclear power plant is available for ground transmitter. Doppler shift requires compensation. Optical links probably will be highly effective when precise tracking can be accomplished.
(8)	Moon Base on Opposite Side from Earth	Earth	Relay Stations are essential. These may be (1) a circumlunar cable to near side base, (2) a recording lunar satellite for readout on near side, (3) a tower-to-tower circumlunar link to near side base, (4) a network of lunar satellite relays phased 120 degrees in longitude (and at highest stable altitude) illuminated by a directive, tracking antenna at far-side base and readout by tracking antennas on Earth, or (5) perhaps the orbiting electron beam link "launched" from a high tower to a high tower electrode at the near-side base (feasibility of this idea is still debatable).

Table 3-2 (continued). Communications subsystem summary table.

	First Terminal	Second Terminal	Problems and System Limitations
(9)	Earth	Lunar Satellite	Conditions similar to (4) above. Satellite is obscured by Moon during nearly half of each orbit except when its orbit plane is parallel to local horizon(s) of Earth station(s). If continuous communication is required, a network of phased satellites may be used as mutual relays in such a way that at least one is always visible to each other one and to the Earth tracking antennas.
(10)	Earth	Planetary Satellite	While planet is not near opposition or conjunction, no unusual solar noise problem arises due to antenna illumination by the direct Sun. Problem is similar to (9) and (4) above except for distance which implies higher power and directivity and a time delay due to two-way transit at velocity of light. Planet can be eclipsed by Sun or Moon and satellite is eclipsed by planet during part of its orbit except under limited, special phasing conditions. Planetary satellite is assumed to be power limited, so directive antenna is desirable; this implies requirement for a steerable antenna, a spatial attitude reference (star trackers), an Earth tracker and sufficient precision of tracking to bring the Earth into the antenna's major lobe. Mariner, Ranger and Voyager studies apply. Bandwidth is severely limited unless power is in multimegawatt region.
(11)	Lunar Roving Vehicle	Earth	Earth tracking directive antenna can be under Earth-terminal's remote control, if an omnidirectional acquisition antenna is provided on the LRV; if the LRV is manned, initial acquisition can be "eye balled" in. If Earth is below LRV's horizon, some relay link is necessary (if contact is to continue) and the problem is similar to (8) above.
(12)	Planetary Roving Vehicle or Colony	Earth	Conditions are similar to (11) above but suffer from much greater range and time lag. Rotation of planet regularly obscures a single planetary base even with total Earth longitude coverage (three or more DSIF stations). Remedy is to provide one (or a set of three or more longitudinally phased) relay satellite(s) around planet to act as relays. Multiple-doppler, refraction, and atmospheric absorption problems exist. Bandwidth is severely limited until nuclear power and large directive transmitters can be installed on planetary surface. Intermittent schedules required when minimal systems are used.
(13)	Spacecraft or Space Station	Spacecraft	Communications limited to line-of-sight, unless relay vehicles are used. High powered, high bandwidth laser links will permit extreme ranges ($> 10^9$ kilometers) if attitude control and directivity permit. Where much more than adequate laser beam power is available, diffusers can serve to relieve pointing requirements. UHF, SHF and occasionally VHF will be used in the near future (for MOL, etc.).
(14)	Outer Planets	Earth	Extremes of power and directivity are required at both terminals. Time lags between call and reply extend to many hours. If SHF or optical links are to be used, a relay station in Earth orbit or (preferably) on the Moon is desirable to permit conversion to UHF before entering Earth's atmosphere. Bandwidth severely limited (unless power is extremely high) for radio. High power and directivity with beam spreading diffusers may be required to permit use of wide band laser links. Eventual X or gamma laser links may permit solving bandwidth/power problems. Early probes may record all data during planetary flyby, then readout during closer approach to Earth; however, this does not solve the command and control problem and would imply delegating all control to an on-board programmer.
(15)	Re-entering Spacecraft	Earth	Spacecraft is "blacked out" by a shield of plasma surrounding it. Recent Gemini system tests indicate that considerable progress has been made in overcoming the problem.

Table 3-3. Planetary fly-by scientific payload.

Interplanetary Measurements			
	Sample Rate	Bits Per Sample	Total Bits
Solar magnetic field	2/hr.	32	1.7×10^7
Solar wind	1/hr.	200	11×10^7
Cosmic dust	1/hr.	100	5.3×10^7
Lyman α line intensity	1/hr.	8	0.42×10^7
X-ray flux	1/hr.	24	1.3×10^7
Cosmic ray flux	3/hr.	72	3.9×10^7
Solar flare proton flux	1/hr.	24	1.3×10^7
			1.5×10^8
Planetary Fly-By Measurements			
Magnetic field	60/hr.	16	10^4
Trapped radiation	66/hr.	56	3×10^4
Atmospheric composition	612/hr.	256	1.6×10^4
Surface features	300/hr.	2.5×10^6	14×10^7
			1.4×10^8

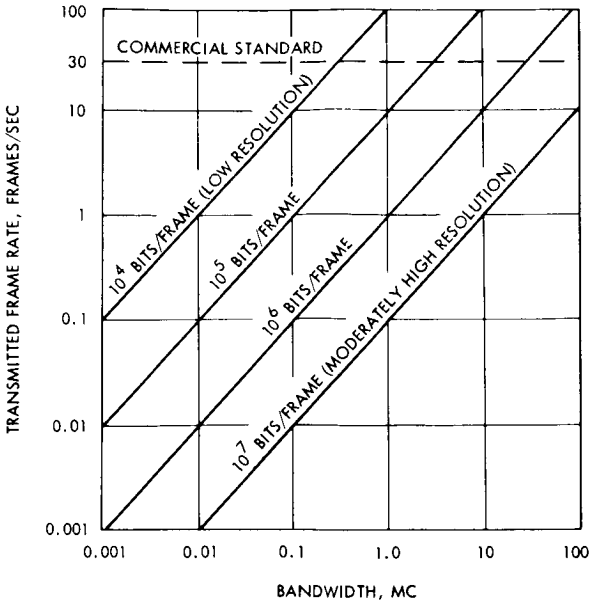
requirements versus bandwidth are shown in Figure 3-6 with resolution as a parameter. Figure 3-7 shows the anticipated data rates for various communication tasks for a 250 million mile range typical of the near planets.[5] The lower shaded horizontal bar indicates the data rate required for transmitting one television picture per day. Anticipated Voyager communication requirements are listed in Table 3-5.[3] Depending on the data rate requirements of a particular mission, it may be preferable from a systems viewpoint to provide sufficient data storage capacity to permit transmission at rates far lower than the acquisition rate. An alternate possibility is transmitting stored data at the real-time rate, but at a later time when the communication range has decreased because of the vehicle trajectory. This may pose a reliability problem.

Table 3-4. Typical data requirements.

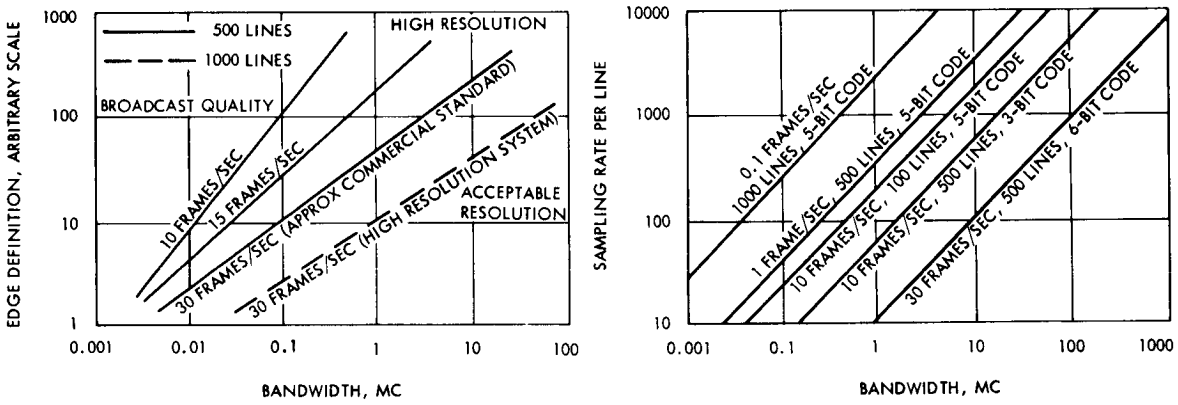
Pulse Code Modulation				
Type	Signal Bandwidth	Representation, bits/sample	Maximum Bit Rate	Tolerable Error Rate
Scientific data	0-1 kc	9	18 kilobits	10^{-2}
Engineering data	0-1 kc	7	14 kilobits	10^{-2}
Command data			50 bits	10^{-5}
Teletype			75 bits	10^{-5}
Speech	0-4 kc	5	40 kilobits	10^{-3}
Real-time television (500 x 500 element picture)	15 cps-4 mc	6	48 megabits	10^{-3}
Pictorial transmission (500 x 500 element picture in 12.5 seconds)	15 cps-10 kc	6	120 kilobits	10^{-3}

Spaceborne recorders are available with speed capabilities ranging from 1 bit/sec to 5 megabits/sec. Recorders with write/read speed ratios of 10^2:1 have been used in space, although higher speed ratios of 10^4 or more would be desirable for some applications. In general, recorder system weight is a direct function of the capacity for a given storage mode (e. g., tape), plus a fixed weight which is a function of the data rate. As an example, the Mariner 4 data tape recorder and its data encoder weighed 39.5 pounds and stored 5.25×10^6 bits of data. Actual operating mean times between failure are on the order of 1,000 hours, although recorders have survived a primarily quiescent existence in the space environment for as long as one year.[5]

Data handling requirements may be as small as 10 to 100 computational cycles per day during midcourse on a planetary fly-by mission, then rise to 100,000 per second during encounter. A typical aerospace computer having an add time of 2 to 5 microseconds and a capacity of



BANDWIDTH REQUIREMENTS CAN BE REDUCED MOST HANDILY BY STRETCHING THE SIGNAL IN TIME (THAT IS, REDUCING THE NUMBER OF FRAMES TRANSMITTED PER SECOND), AS THIS GRAPH SHOWS. HOWEVER, THE NATURE OF THE ANALOG SIGNAL PUTS AN EFFECTIVE LOWER LIMIT ON FRAME RATE.



REAL-TIME TRANSMISSION OF NARROW-BANDWIDTH SIGNALS IMPOSES LIMITATIONS ON BOTH ANALOG AND DIGITAL SYSTEMS. IN ANALOG (RIGHT), IT RESULTS IN PICTURES OF LOW CONTRAST, INDICATED HERE IN TERMS OF EDGE DEFINITION. IN DIGITAL (LEFT), IT FORCES A LOW SAMPLING RATE.

Figure 3-6. Television bandwidth requirements.

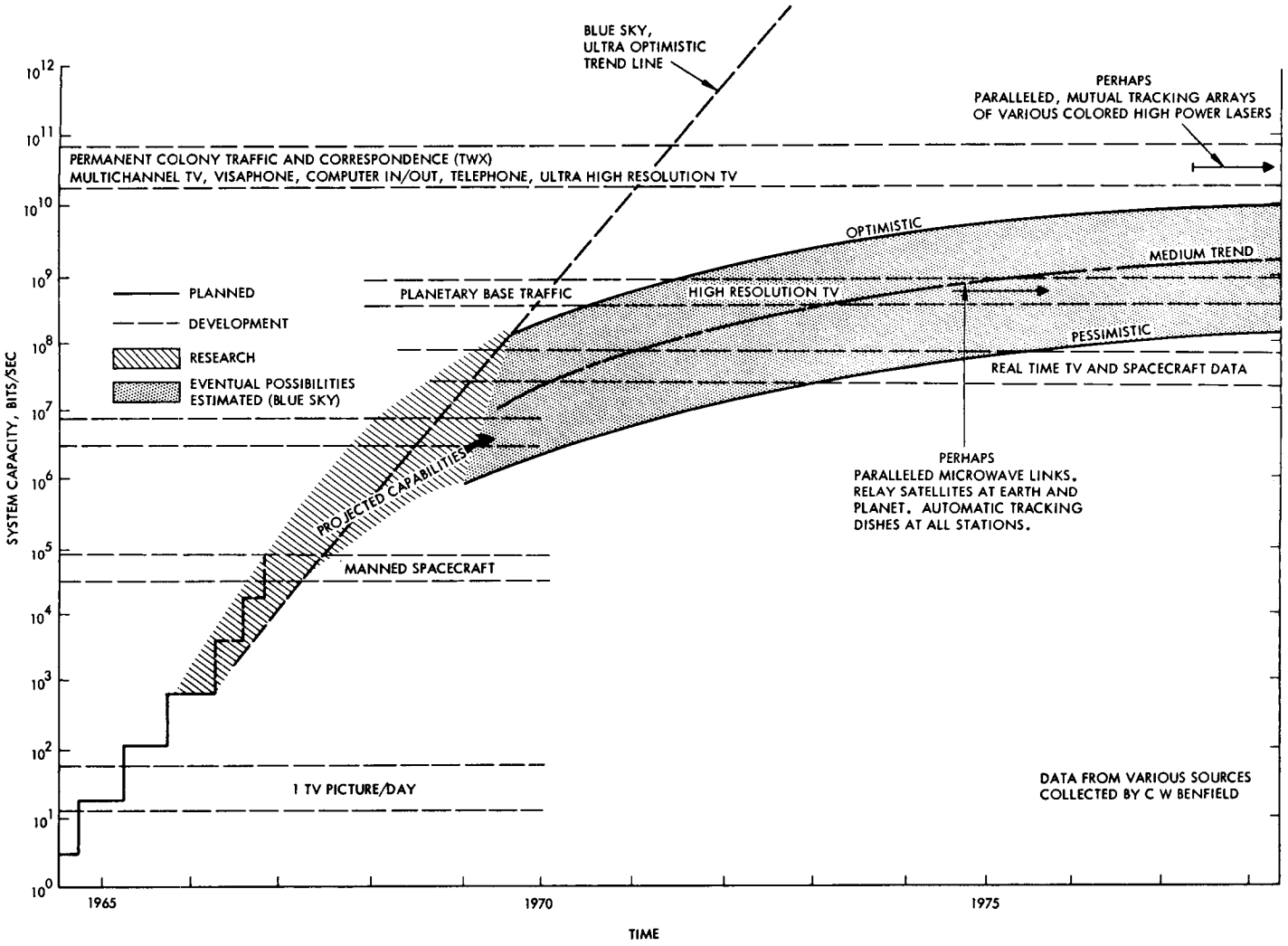


Figure 3-7. Projected communication requirements for near planets. Message time, or data rate, can be improved by technological advances. Data rate appears as bits/second for a typical planetary distance of 240 million miles, with mission capabilities represented by various bit rates.

Table 3-5. Voyager spacecraft preliminary designs

	Boeing	GE	TRW
SATURN IB LAUNCH			
High gain antenna size	8 ft x 12 ft	7.5 ft circular	6.5 ft x 5.5 ft
Data storage	2×10^8 bits	6×10^8 bits	2×10^8 bits
Solar panel area	241 sq ft	197 sq ft	190 sq ft
Battery power	2460 watt-hr	2280 watt-hr	2000 watt-hr
Engine	solid	liquid	solid
Relay link power and frequency	14 watts/ 100 mc	20 watts/ 200 mc	20 watts/ 137 mc
S-band encounter data rate	8000 bps	8500 bps	5000 bps
SATURN V LAUNCH			
High gain antenna size	8 ft x 12 ft	7.5 ft circular	9.5 ft circular
Data storage	3.8×10^8 bits	6×10^8 bits	2.3×10^8 bits
Solar panel area	316 sq ft	181 sq ft	290 sq ft
Battery power	2720 watt-hr	3240 watt-hr	1890 watt-hr
Engine	solid	solid	liquid
Relay link power and frequency	14 watts/ 100 mc	20 watts/ 200 mc	20 watts/ 137 mc
S-band encounter data rate	15,000 bps	15,000 bps	15,000 bps

Mariner's disquieting evidence of a 5-10-mb Martian atmosphere forced an upward revision of Voyager's launch-weight requirements. Contractors were invited to modify their previous recommendations (first set of data in the table) to take advantage of a Saturn-V-boosted payload.

4096 16-bit words, occupies about $0.05 m^3$, weighs about 10 kg and consumes 80 watts. Compact computers with capacities of over 16,000 36-bit words each are within the present state-of-the-art. It is expected that space computers of the mid-1970's, having a memory capacity of 32,000 36-bit words, would weigh less than 1 kg and displace

less than 5,000 cc's, based on the present promises of thin film and integrated circuitry electronics. Mean times between failure of present computers are on the order of 10,000 hours; mean times between failure on the order of 50,000 to one million hours are predicted by the mid-1970's.[5]

3.3.2 Communication Range

If all other factors remain constant, the information capacity of a communication link decreases as the square of the transmission distance. Thus, the Mars-to-Earth transmission capacity of present facilities is of the order of 10^{-6} that of the Moon-to-Earth capacity. Maintaining a given data rate transmission capability over increased range requires increases in transmitter-power, transmitter-aperture, receiving aperature, or some combination of these. Communication range for a given space mission depends on the launch date and the injection energy expended as well as the objective. Figures 3-8 through 3-11[8] show communication distance vs. launch date with C_3, the injection energy of the escape hyperbola, as a parameter for Mercury, Venus, Mars, and Jupiter missions. The terms Type I and Type II refer to the two possible elliptical interplanetary transfer orbits. For Type I, the heliocentric central transfer angle is less than 180° and for Type II it is greater than 180°. Class I or II refers, to planetary encounter at the first or second intersection of the spacecraft trajectory with the planetary orbit.

3.3.3 Communication System Lifetime Requirements

System lifetime requirements influence the design of a communication system primarily through the imposition of a reliability constraint. System lifetime is determined by flight time and by mission objectives. Flight time versus launch date are given in Figures 3-12 through 3-15[8] for the interplanetary missions of Figures 3-8 to 3-11, with planetary escape hyperbola injection energy, C_3, as a parameter. Further mission life beyond flight time may vary from hours in the case of a planetary fly-by to years in the case of a communication

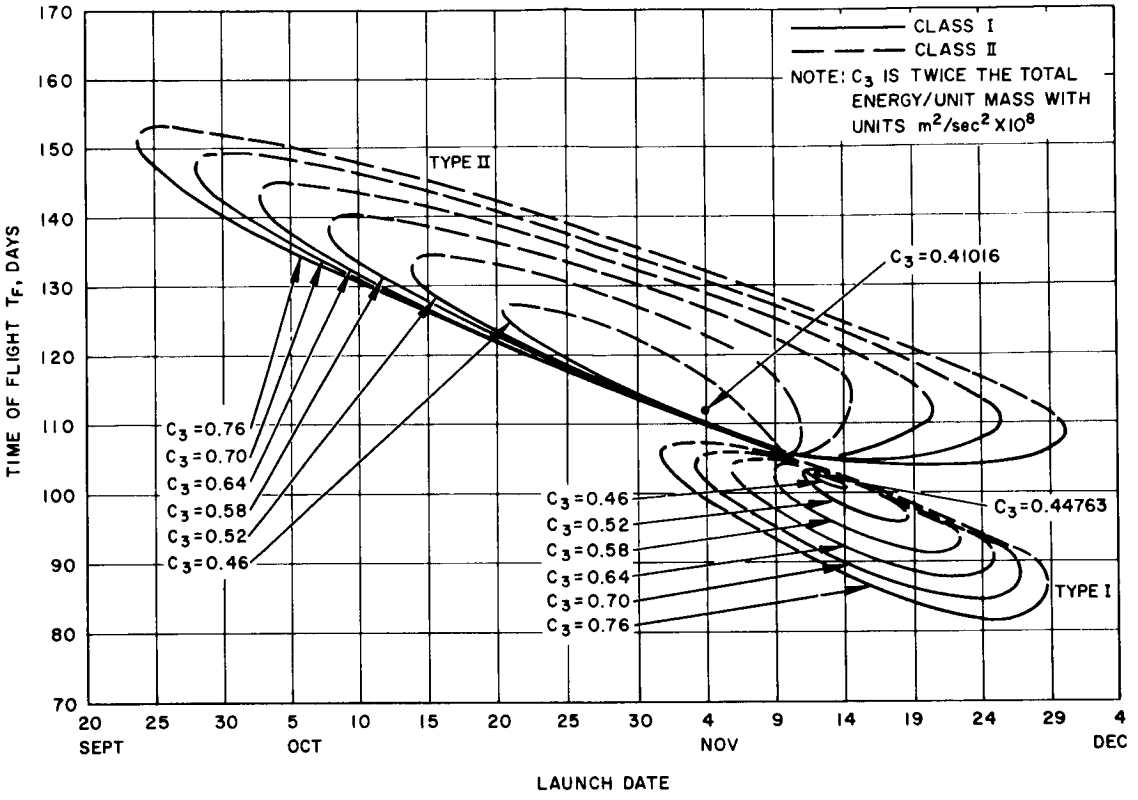


Figure 3-8. Mercury 1968: time of flight versus launch date.

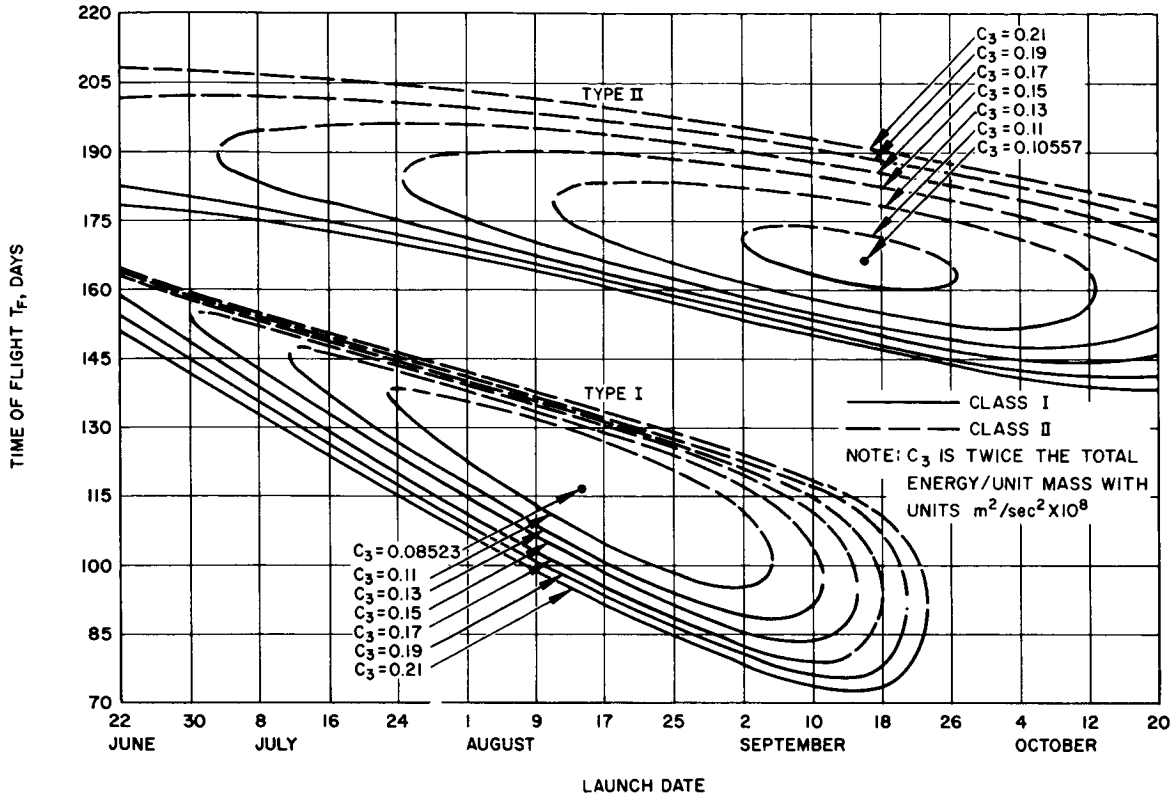


Figure 3-9. Venus 1970: time of flight versus launch date.

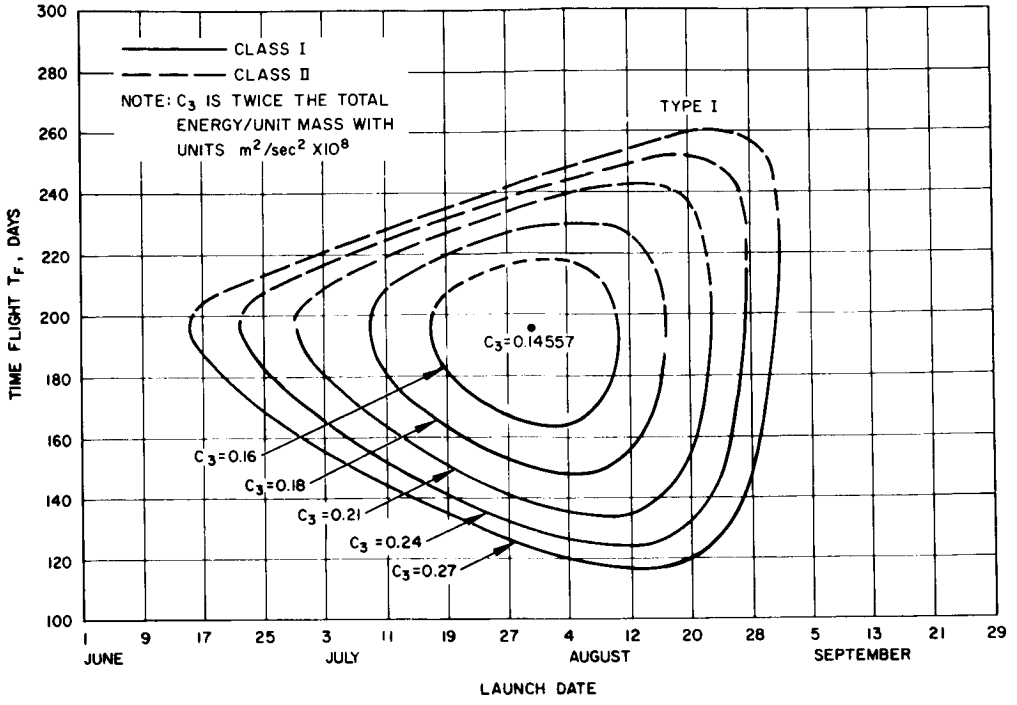


Figure 3-10. Mars 1973: time of flight versus launch date.

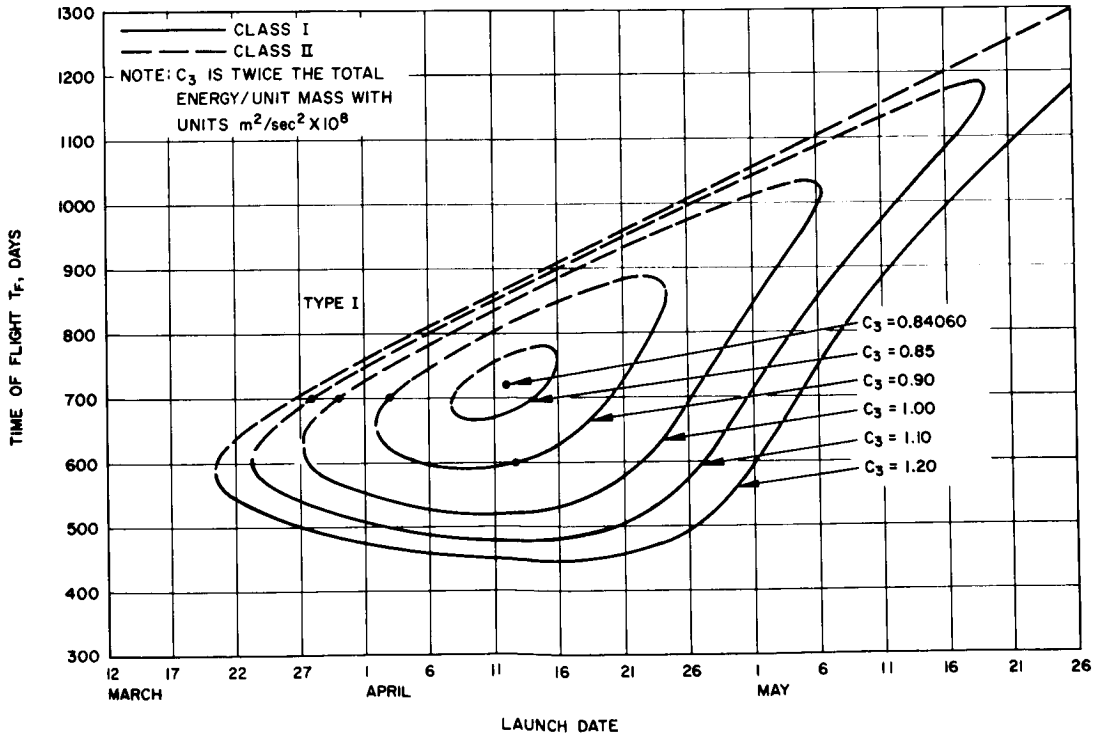


Figure 3-11. Jupiter 1973: time of flight versus launch date, Type I.

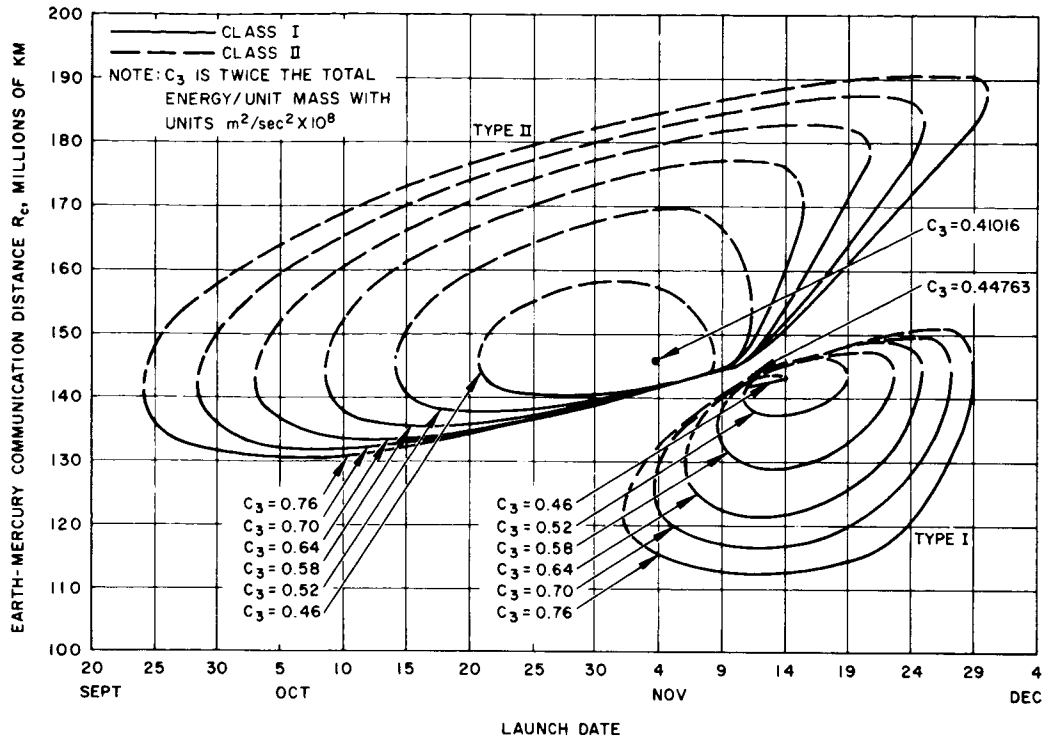


Figure 3-12. Mercury 1968: Earth-Mercury communication distance versus launch date.

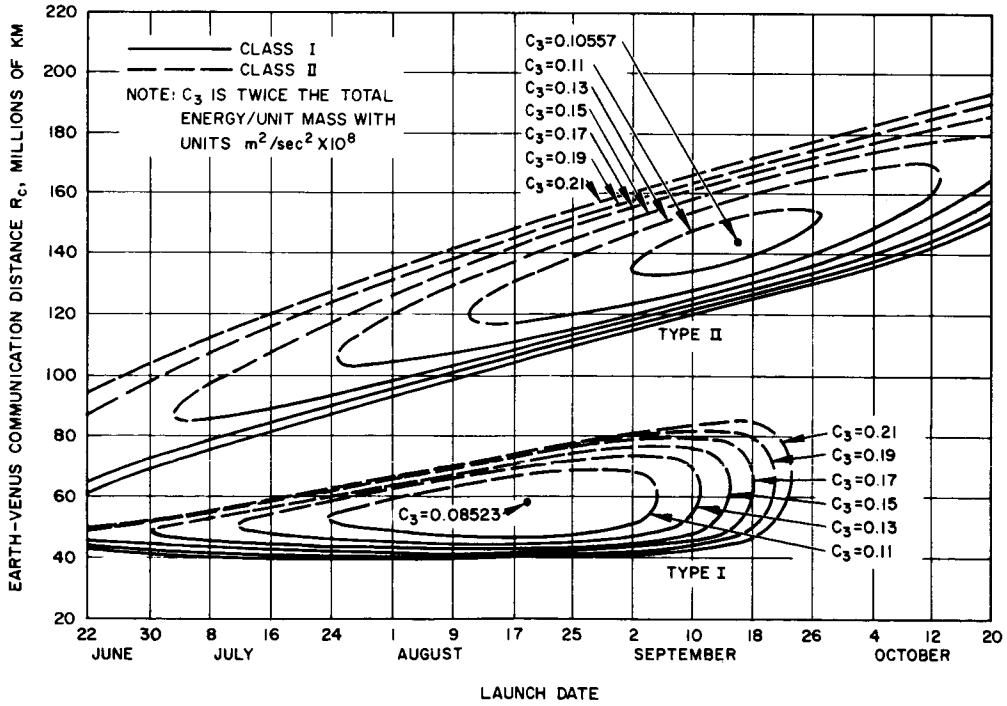


Figure 3-13. Venus 1970: Earth-Venus communication distance versus launch date.

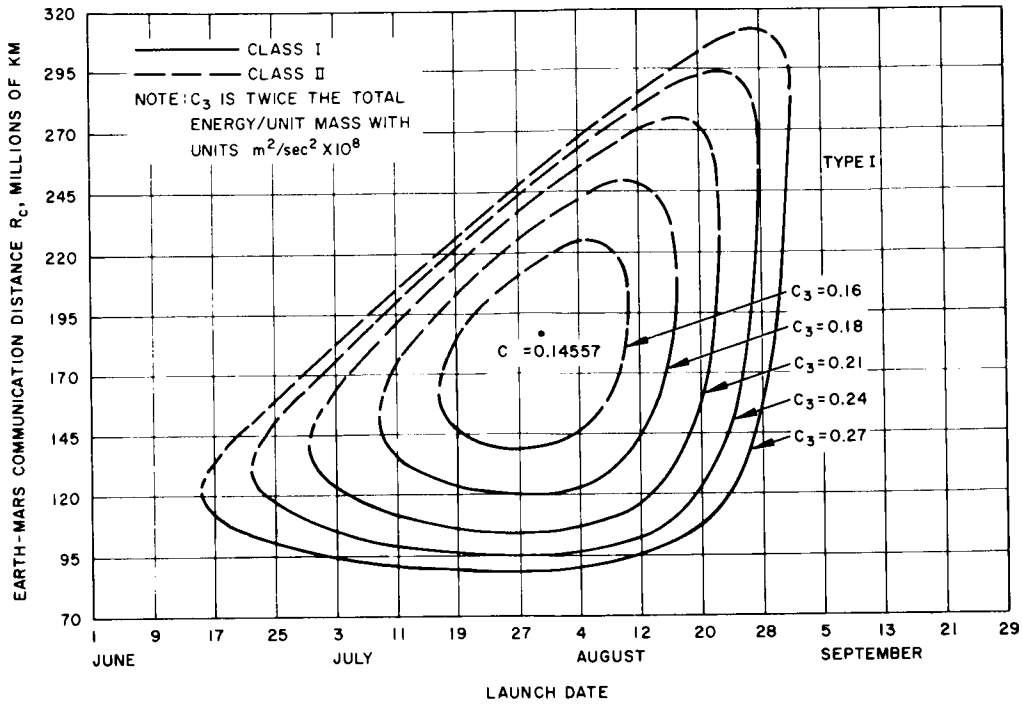


Figure 3-14. Mars 1973: Earth-Mars communication distance versus launch date.

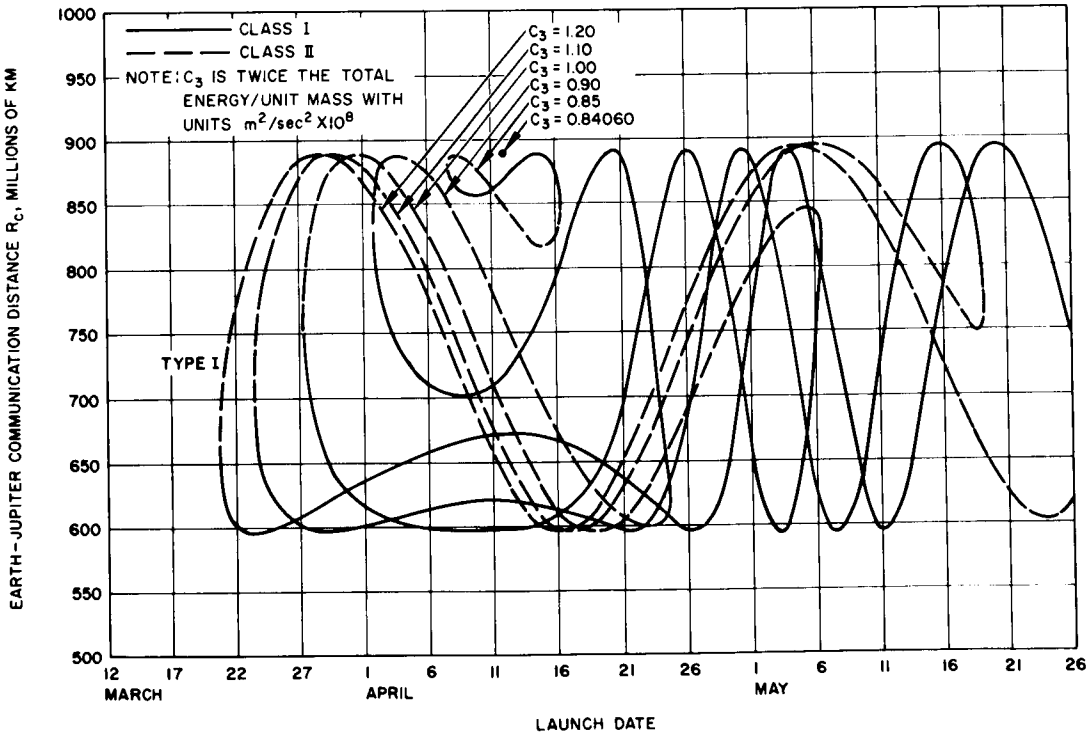


Figure 3-15. Jupiter 1973: Earth-Jupiter communication distance versus launch date, Type I.

satellite. Table 3-6 summarizes some of the projected durations of various types of missions.

3.3.4 Communication System Weight Restrictions

Permissible payload for a given mission is dictated by launch vehicle capability; allocation of that payload is determined by mission objectives. The fractional part of the payload comprised by the communication system will vary depending on the required data rates and transmission range. Payload capabilities of present and projected launch vehicles are depicted in Figure 3-16.[9] It may be seen from Figure 3-16 that the payload weight which may be launched on a given mission (i. e., at a specified characteristic velocity) is a discrete rather than a continuous function. Hence vehicles having weights intermediate between the payload capabilities of two launch vehicles may be increased in weight to the payload of the next largest launcher without penalty. Thus for some payload weights, additional weight

Table 3-6. Typical mission durations.

Mission	Time Period	Probable Duration
Orbital:	Manned Missions:	
Apollo Applications	1968-1972	45 days
Manned Orbiting Research Laboratory	1973	1-5 years
Lunar:		
Apollo Applications	1971	14 days
Semi-permanent base	1976-1980	6 months
Planetary:		
Mars or Venus Fly-by	1976-1980	15-20 months (typical opposition class mission)
Planetary:	Unmanned Missions:	
Voyager	1973-1975	8-12 months
Deep Space:		
Advanced Pioneer	Post 1972	Several years

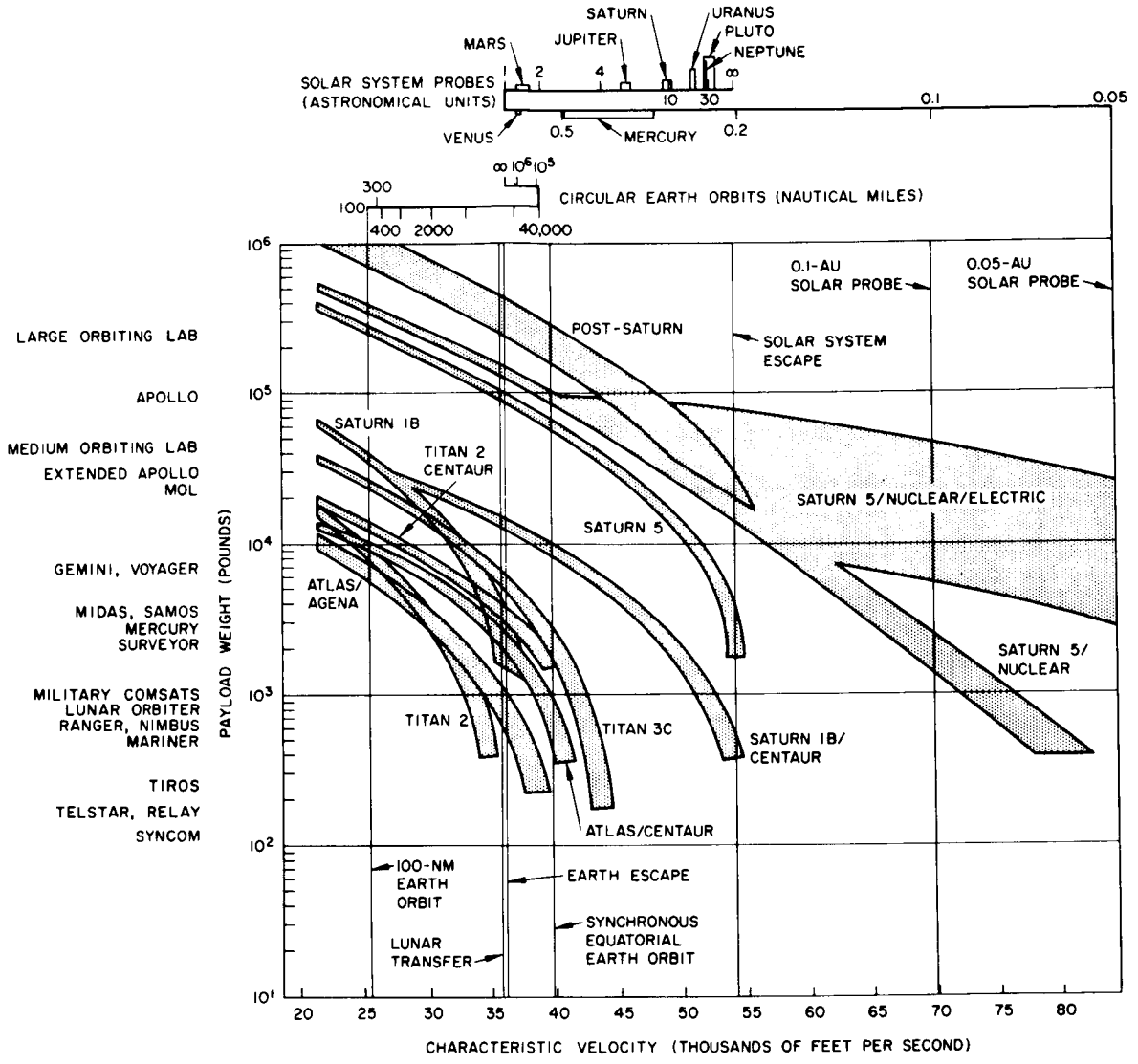


Figure 3-16. Launch vehicle capabilities[9] - payloads and destination. (The horizontal characteristic velocity scale shows velocity a vehicle needs at 100 n. mi. altitude (not orbit) to reach destination, assuming a minimum energy trajectory.

may be a non-critical burden. An example of such a payload weight quantization is the rather wide gap in payload capability between Saturn IB/Centaur and Saturn V.

3.4 CONCLUSION

The determination of the minimum fraction of total payload weight to be allocated to the spacecraft communication system for each category of mission envisioned will be the subject of extensive consideration in succeeding phases of this program. The probable instrumentation and the corresponding data acquisition rates, the communication requirements, and the transmission ranges for unmanned and manned orbital, lunar, and interplanetary missions will be surveyed. Once the communication system capability required for a particular mission has been established, along with any limiting constraint, the optimum division of allotted communication system burden among the various system functions for minimum weight or cost will be ascertained. This optimization will take into account trade-offs between such interrelated system parameters as:

a. data transmission rate and information storage capacity.
b. data transmission rate and system reliability.
c. information storage rate and storage utilization factor.
d. aperture size and tracking accuracy requirement.

The resultant systems will uniquely fulfill the communication requirements determined for a given mission. These systems must then be related to the ground telecommunications and tracking networks to ascertain whether adequate ground instrumentation and operations exist to perform these missions.

3.5 REFERENCES

1. Morgenthaler, G. W. and Fosdick, G. E.,"Selection of Launch Vehicles, Spacecraft, and Missions for Exploration of the Solar System,"Research Report R-64-6, Martin Company, 1964.

2. Summary Report, Future Programs Task Group, National Aeronautics and Space Administration, Washington, D. C., 1965.

3. Stambler, Irwin, "Mariner/Voyager: A New Fix on Mars", Space/Aeronautics, Volume 45, Number 5, May 1966, p. 68.

4. "Manner Space Flight-Beyond Apollo", Space/Aeronautics, Volume 44, Number 5, October 1965, p. 52.

5. Benfield, Charles W., NASA Pattern Relevance Guide, Volume III, Honeywell, Military Products Group, October 1965.

6. Private Communication, Elwyn Angle, Member of the Technical Staff, Hughes Aircraft Company, October 1965.

7. Leary, Frank, "Television from Space", Space/Aeronautics, Volume 41, Number 3, March 1964, p. 70.

8. Design Parameters for Ballistic Interplanetary Trajectories, Jet Propulsion Laboratories Technical Report No. 32-77; Part I, "One Way Transfers to Mars and Venus," January 16, 1963 and Part II, "One Way Transfers to Mercury and Jupiter", January 15, 1966.

9. Roman, Paul, "The Booster Crossroads", Space/Aeronautics, Volume 43, Number 3, March 1965, p. 58.

5.0 TRANSMITTING POWER SOURCES

5.1 INTRODUCTION

This section presents a discussion of rf and optical power sources. The division of the spectrum for these two areas was arbitrarily chosen at approximately 300 microns. These rf sources include microwave, millimeter wave and submillimeter sources and optical includes uv, visible and infrared (to 300μ).

The approach taken has been to include in each section some of the theory of operation of these devices as well as present performance values. This theory can then provide the basis for interpretation and (possibly) extrapolation of the values given. In addition it can display those constraints on parameter values which limit their use to mutually compatible regions of parameter space.

In addition to the listing of the present state-of-the-art performance, salient trade-off in terms of "measures of merit" for the various devices are given. Parameter vs burden relationships are given when possible, but should at this point be considered as preliminary.

5.2 RADIO FREQUENCY SOURCES

It is the purpose of this section to review the theoretical design considerations of known radio frequency sources. The review is limited to those factors which predict the principle electrical and mechanical characteristics which affect the use of the source in tracking and communication systems. In this sense, sources are meant to include both oscillators and amplifiers since either may represent the basic transmitter device. A natural division results as a function of frequency because of the fundamental mechanisms involved in the conversion of dc to rf energy. For the present purposes, the frequency range is divided and defined as follows.

uhf	100 mc	to	1000 mc
Microwave	1 Gc	to	30 Gc
Millimeter	30 Gc	to	300 Gc
Submillimeter	300 Gc	to	1000 Gc

In addition, the discussions which follow are limited to cw devices.

5.2.1 Radio Frequency Source Theory

UHF Sources. In general, negative grid tubes dominate the uhf frequency region. Conventional tubes such as triodes and tetrodes are utilized in external resonant circuits for Class A or B operation. At the higher frequencies these resonant circuits take the form of coaxial lines. Electron transit time effects limit the extension of these techniques to higher frequencies and in this area, power limitations occur due to the thermal capability of the collector and the envelope seals. The relative advantages and disadvantages of this class of device is summarized in the following Table 5-1.

Microwave Sources. At frequencies where electron transit time effects are significant, some form of velocity modulation tube is employed to overcome these limitations. Basically, this technique provides a means for a bunch of electrons containing the input dc power to remain in approximate time or space synchronism with a component of the ac wave so power can be transferred continuously. Transit time effects are no longer of consequence because an electron continues to see the same phase of the ac field in spite of the axial motion of the bunch of electrons. The electrons give up either kinetic energy in the case of a klystron or traveling wave tube, or both kinetic and potential energy in a crossed field tube. The interaction region normally takes the form of a slow-wave circuit in the vicinity of the electron bunches such that the electrons see the rf fields associated with the wave propagating structure. The various types of these tubes are summarized in Table 5-2.

Advantages	Disadvantages
1. Good phase stability and tracking characteristics, because of short transit time. Typical sensitivity to change in voltages is: Screen – 1 degree per 1 percent ΔE_{sg} at 1 gc Plate – 0.5 degree per 1 percent ΔE_p. Plase variations due to drive level and filament voltage changes are negligible. Because the gridded tube has short electrical length, its phase tracking characteristics are determined primarily by the associated circuitry. Mass-produced double-tuned circuits show typical phase track deviations of 5 degrees over the usable bandwidth. A suitable secondary coupling can be chosen to give linear phase-frequency characteristics.	1. Limited gain-bandwidth product. 1,000 to 2,500 mc for class A operation and 500 to 1,000 mc for class B are typical. These values hold for operations to about 500 mc; above this frequency, the product drops off rapidly.
2. High Efficiency. At 500 mc, typical plate efficiencies approach 70 percent for class C operation. Overall efficiencies of 50 percent, including filaments, are common.	2. Operating frequency is limited to less than 1,000 mc. Above this, transit time is long enough to create current wave-form distortion in the plate current. The result is decreased gain-bandwidth product and lower efficiency.
3. Economy. Most coaxial tubes have a simple structure and are easily mass-produced. It is usually possible to select a tube already in mass production. This affords additional economy.	3. Higher-power tubes are disproportionately expensive. This is attributed to the fact that higher-power tubes have not been manufactured in such large quantities as the smaller ones.
4. Inherent Filtering Action. For a double-tuned reentrant cavity, in which a tetrode normally operates, the roll-off is at a rate of 12 db per bandwidth octave. For a 2 percent bandwidth device at 500 mc, typical second harmonic suppression is more than 80 db for class B operation.	4. Reliability and life expectancy. The cathode of gridded tubes must deliver a greater current density than that required by other devices. In addition, maximum r-f current must be instantaneously available for the cathode, thus requiring that the cathode operate at a comparatively high temperature. As a result, normal cathode life expectancy is generally 2,000 to 5,000 hours, though some tubes are rated at 5,000 to 15,000 hours. However, arcing and other failure mechanisms reduce the MT of most types to about 3,000 hours.
5. Much more experience exists in design of vacuum tubes than in other types.	5. The tetrode is a filamentary device requiring a variety of electrode voltages. Distribution of these voltages within a complex system can be troublesome.
6. No focusing magnets or field are required, therefore weight and volume are reduced.	6. Arcing problems are generally greater than in other devices due to close spacing of elements.
7. Tetrodes can operate from a dc supply on the plate. Because the control or screen grid can act as a switching element, efficiency is greater and noise is lower.	7. High output capacitance limits bandwidth.
8. Tetrodes are constant-current generators and tolerate mismatches better than devices depending upon traveling- or standing-wave operation.	
9. They are relatively insensitive to temperature up to the rating of the envelope seals.	

Table 5-1. Negative grid tube characteristics

			Interaction Circuit					
			Backward Wave		Forward Wave		Standing Wave, cavity	
			Oscillator	Amplifier	Oscillator	Amplifier	Oscillator	Amplifier
Electric-Magnetic Fields	Crossed Field (M-type)	Injected Beam	M-BWO M-Carcinotron	M-BWA Bitermitron	TPOM	CFA TPOM Bimatron		
		Continuous Cathode	Stabilotron	Amplitron CFA	VTM	FWA-CFA Dematron	Magnetron	Circlotron
	Linear (Single or Parallel) Field (O-type)	Injected Beam	O-BWO O-Carcinotron	O-BWO		TWT TPO	Klystron Reflex Klystron Monofier Monotron	Klystron

Table 5-2. Type classification of microwave tubes — oscillators and amplifiers

The important types of tubes shown above may be reviewed as follows.

Klystrons. The klystron amplifier is a well developed, reliable, and in many cases a long-life device. If conservatively designed, it should be capable of meeting space requirements. The requirement for a high voltage, high power modulation technique reduces the overall efficiency of a transmitter chain using a klystron as the final power amplifier.

Table 5-3 summarizes the significant advantages and disadvantages of klystrons.

TWT's. The traveling wave tube is inherently a high average power amplifier so there exists no problem in achieving the parameters necessary in a space system. Furthermore, the twt can be made with arbitrarily large gains without sacrificing any of its outstanding electrical characteristics, and without increasing prohibitively the overall package size and weight. Periodic focusing of twt's has resulted in a very light weight, compact structure ideally suited to spaceborne applications. Herein lies one of the significant advantages

Advantages	Disadvantages
1. Single envelope is practical for gains up to 30 db, depending on output power levels.	1. Longer electrical length per db of gain as compared to cross-field devices.
2. Much development experience exists both at high power (megawatts) and lower drive levels (tens of kws).	2. Gains are about half those obtained with TWTs.
3. Bandwidths to 10 percent have been achieved in high-power units; 5 percent or less is more realistic at low power.	3. Klystrons require a filamentary cathode and gun for operation, thus more electrode voltages are needed.
4. Klystrons can be made to operate with dc beam supplies by employing switching or modulating anodes.	4. Higher beam voltages required for a given output power than in crossed-field devices.
5. Klystrons still offer higher power than any other tube type.	5. Bandwidth limitations are severe at low power (tens kws); 2 to 5 percent is typical.
6. Focusing is normally performed by solenoids.	6. Good voltage regulation may be required for acceptable phase stability.
7. More design and development experience exists on this type of beam device than on any other.	7. Efficiencies of 25 to 35 percent are typical.
8. Higher perveances have been achieved than in TWTs. This permits lower beam voltages for equivalent output power.	8. Structure offers moderate filtering; may require separate high power filters at an additional cost to the system.
9. System reliability is enhanced in cases where 30 db gain is sufficient, since a single amplifier stage is adequate.	9. Reliability not reasoned to be as good as cold cathode devices operated without gun.
10. Cooling techniques are known and optional; air and liquid are common.	
11. Since the klystron is a unidirectional device, operation into a mismatched load is possible without isolation, depending on system limits of power and phase shift.	
12. More recently developed electrostatically focused klystrons offer reduced size and weight.	
13. Shorter electrical length per unit gain than the TWT, thus suffering less voltage-phase sensitivity.	

Table 5-3. Klystron characteristics.

that the twt possesses compared to its counterpart in the klystron field. At present, klystrons are focused with heavy, bulky permanent magnets of the horseshoe variety. These large magnets create an extensive leakage field which affects all of the surrounding electronics, and furthermore, there is no easy method of shielding these fields without degrading the klystron performance.

By appropriately designing the electron gun optics so that the emission density at the cathode surface is quite low, twt's can be made to yield an arbitrarily long life. This has been established, for example, with the Bell Telephone Laboratory twt designated for the transcontinental radio relay link. Hundreds of these devices have been under continuous life test for more than four years without a single

failure attributable to the basic tube design. An expected life greater than 40,000 hours has been established beyond all reasonable doubts for these tubes.

Table 5-4 summarizes the significant advantages and disadvantages of traveling wave tubes.

Cross-Field Devices. Under this general category falls the conventional magnetronoscillator, the amplitron amplifier, and the many linear beam-type magnetron amplifiers. This class of tubes is in general the most efficient and are lighter and smaller per kw of output power than any other device. The basic problem with these tubes is life. Recent advances such as the coaxial magnetron oscillator show promise of increasing the life; however, extensive data is not available at this time. Table 5-5 summarizes the significant advantages and disadvantages of cross-field devices.

Advantages	Disadvantages
1. Single envelope is practical for gains to about 60 db.	1. Long electrical length per db of gain in comparison to other tubes.
2. Much development experience exists both at high power levels (megawatts) and lower drive levels.	2. Phase sensitivity of rf output is very dependent on beam voltage, because the device operates on the principle of synchronism between electron velocity and rf velocity on the slow-wave structure.
3. Bandwidths beyond 10 percent are common in today's twt's.	3. Twt's require a filamentary cathode and gun for operation – more electrode voltages are thus needed.
4. Can operate with dc beam supplies by employing switching or modulating anodes. Grid control is possible at power levels below 10 kw.	4. Low perveance of twt's requires higher beam voltages for a given power output compared to other devices. High perveances involve hollow beams, whose increased current density may cause focusing difficulties and cathode loading problems.
5. Long length, small-diameter form factor is suitable for phased arrays limited to tube diameters of less than $\lambda/2$.	5. High gains usually require solenoid focusing (or a recently proposed magnetic matrix focusing technique). Solenoids for high power tubes are large, require substantial power, and create packaging and distribution problems.
6. Several types of slow-wave structures can be cascaded in a single envelope to optimize the rf coupling design	6. Acceptable phase stability calls for good beam voltage regulation.
7. Focusing can be accomplished by electromagnets, permanent magnets, or electrostatically, depending on system requirements	7. Efficiency is $\approx$ 25 percent without depressed collectors. Depressed collectors raise efficiency by $\approx$ 10 percent.
8. Reliability is enhanced by single-stage operation.	8. Electrical structure offers very little filtering. Separate high-power filters may be needed.
9. Depressed-collector techniques ease regulation requirements on high-current beam supply.	
10. Cooling techniques are known and operational; air and liquid are common.	
11. Severed or attenuated slow-wave structures enable operation without circulators into a mismatched load, depending on the limits of phase tolerance.	
12. No doubt exists about development and mass production of twt's.	

Table 5-4. Traveling wave tube characteristics

Advantages	Disadvantages
1. Gains to 45 db possible with injected beam variety; 10 percent bandwidth typical.	1. Limited gain of 10 to 15 db for distributed emission and types.
2. Good phase stability; fairly independent of anode voltages, since velocity synchronism is not an operating requisite.	2. Has very limited dynamic range of output power for a particular anode voltage and varying input power.
3. Distributed emission type has shortest electrical length per unit gain of all tubes.	3. Crossed-field tubes are bidirectional. Any reflections into the output will return directly to the driver or input circuit. A circulator or isolator is thus necessary.
4. Can operate with a cold cathode; electron gun or filaments are avoided.	4. Injected beam types require very good regulation of sole voltage for good phase stability. Typical values are 20 degrees for a 1 percent change in sole voltage.
5. Reasoned to have long life compared to filamentary devices.	5. In distributed emission types it is difficult to initiate emission.
6. Small in size and relatively light in weight.	6. Perveance of injected beam types is comparable to that of twt's, thus requiring higher beam voltages for a given output power.
7. Distributed-emission type provides the lowest anode voltage for a given power output. This simplifies the power distribution system.	
8. Air or liquid cooling is acceptable.	
9. Highest microwave tube efficiencies: Injected beam – 35 to 50 percent Distributed emission – 45 to 65 percent	
10. Operated at saturation to give a constant output independent of input power variation; makes a good limiter.	
11. Can be used in a duplexed system with receiver and duplexer on the low input side.	
12. Structure is ideally suited for mass production under tight mechanical tolerance control.	

Table 5-5. Crossed-field tube characteristics

Millimeter Sources. Most of the successful attempts to develop sources in the millimeter wave region of the spectrum have involved extensions and extrapolations of the microwave device techniques. However, since the slow-wave circuit must have dimensions comparable to a wavelength, the problems of electron control and circuit thermal dissipations become formidable. The linear beam, or O-type distributed interaction device overcomes these difficulties better than the klystron or cross field tube for rather fundamental reasons.

In the traveling wave type of tube, in the ideal limit the electron stream need not touch the rf circuit and the beam collection function is accomplished by a separate and easily cooled electron. A distributed interaction device such as a traveling wave tube has an additional advantage over a single output gap tube such as a klystron. When all

the power must be transferred to the output circuit via a single gap, high Q circuits are involved with the problems of multipactor effects and voltage breakdown.

Surprisingly, the efficiency of the overall device in millimeter wavelengths does not suffer appreciably when compared to microwave tubes. Because the millimeter interaction efficiency is necessarily low, the beam has a more uniform distribution of velocities. Accordingly, the collector electrode can be operated closer to cathode potential without danger of returning slow electrons into the circuit region. This operation results in a recovery of the overall conversion efficiency.

Submillimeter Sources. The conventional techniques of rf device design become thoroughly impractical at frequencies over 300 Gc. Barring a technological breakthrough, one of the better means of producing power at these frequencies involves harmonic generation. Because large powers are available in the millimeter wave region, comparatively low harmonics are needed with reasonable conversion losses. The problem resolves itself to the development of the nonlinear device which can accept large input powers. One of the likely candidates for this application is high pressure gas discharge plasmas. Notable success has been achieved in this area and there appears to be considerable promise for further development.

Scaling Laws. Because of the complexity of the tube design problem it is difficult to reduce the available tradeoffs to simple scaling laws for all parameters. The general rule to indicate the difficulty in achieving a particular performance level related any two designs according to the factor of power times the square of the frequency. While this guide represents a gross oversimplification, it is still probably the best first approximation to a general scaling factor.

Some of the scaling laws pertaining to particular tube parameters can be given crude approximations. Since power, weight, voltage, bandwidth and frequency are most important considerations for system application, the following guide is intended to relate these parameters.

Constant_1	Parameter_2	α	Parameter_3
Frequency	Power	α	$(\text{Voltage})^{1/2}$
Power	Frequency	α	Voltage
Frequency	Weight	α	(Bandwidth%)

Figure 5-1 indicates the frequency power relationship based on an rf heating limitation, while Figure 5-2 indicates the required fabrication tolerances to achieve satisfactory twt performance.

5.2.2 Available Source Performance

Vacuum Tube Sources

Uhf Sources. The present state-of-the-art in power grid tubes varies from a 100-mc tube that can supply 100 kilowatts and weighs 100 pounds to a 1,000-mc tube that can supply 100 watts and weighs 5 pounds.

These tubes have the advantage of requiring no additional magnets for tube operation.

Microwave Sources. Microwave sources fall into two general classifications. First are the relatively low power tubes that are very light and designed for extremely long and reliable life times. Several of these types of tubes have been used in current spacecraft with excellent results. In general, they provide from 2 to 20 watts in the range of 1 to 10 Gc and weigh about one pound. However, these tubes cannot compete on a power/weight basis with the brute force power tubes which have been built without real concern for the total weight of the system. These tubes can demonstrate power in excess of 100 kw although the weight of the tube and magnet system can easily exceed 1,000 pounds. Indeed, current development is in progress on a tube to produce one megawatt of cw power at X-band. The weight and voltage penalty for these tubes is, of course, very high. But for the transmission of the maximum power levels they are without competition.

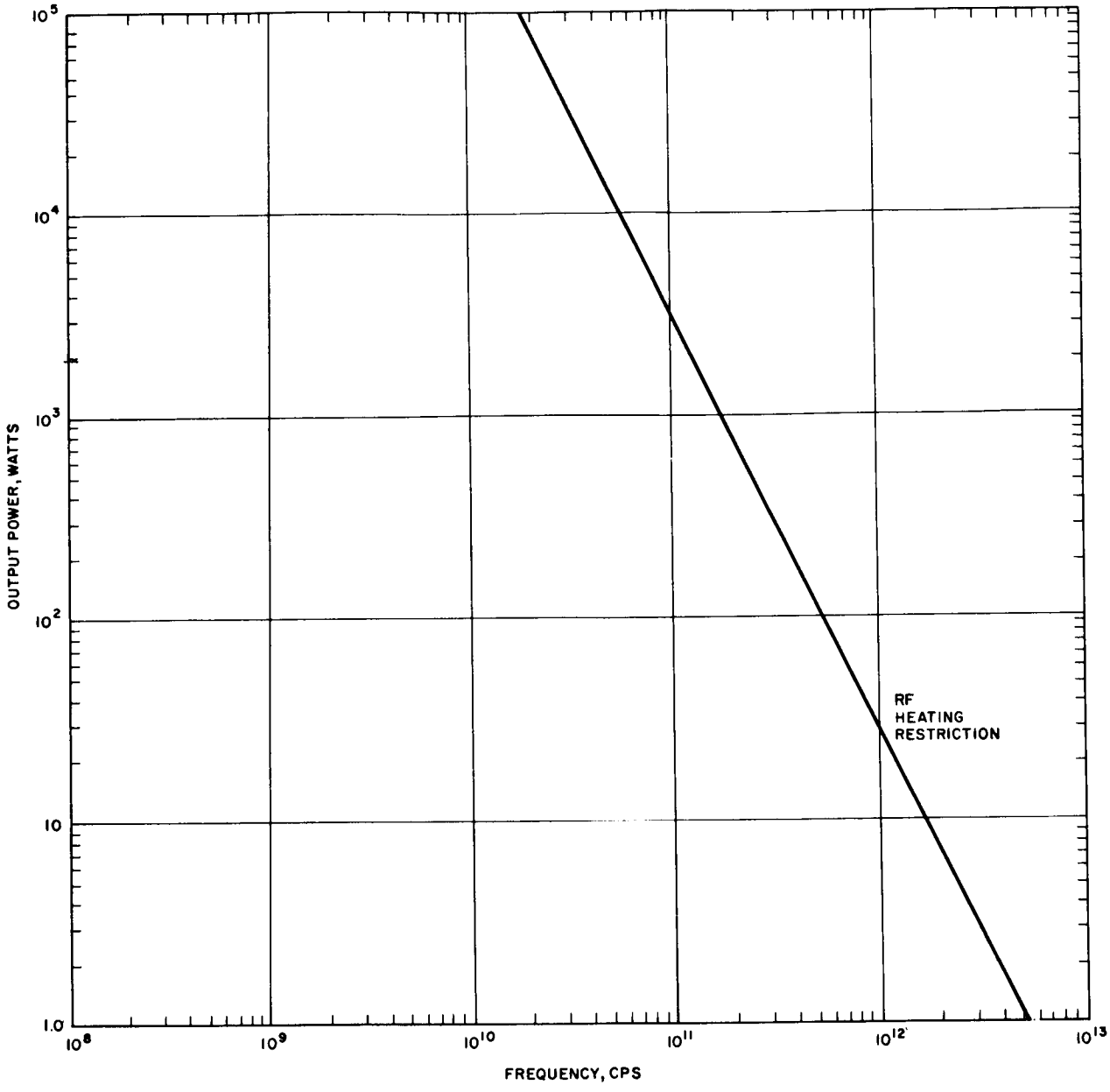


Figure 5-1. Power limitations for a single TWT.

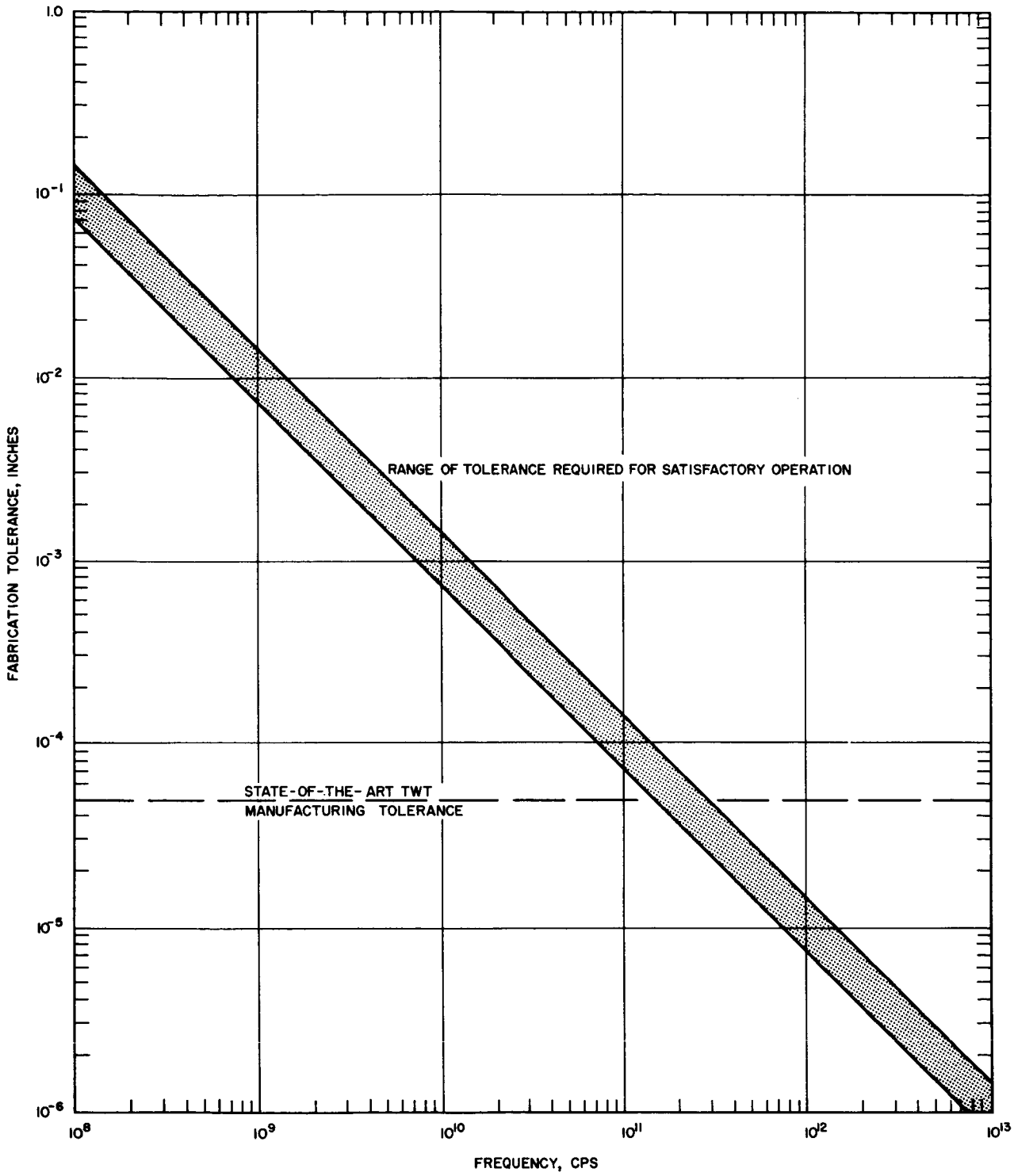


Figure 5-2. Required manufacturing tolerances in TWT as a function of frequency.

Millimeter Sources. Available power levels for mm waves between 30 and 100 Gc have increased by three orders of magnitude since 1960, and efficiency of sources has been increased to be competitive with microwave sources. One kilowatt cw sources are now available at 35 and 55 Gc with efficiencies up to 35 percent. These levels are being achieved in devices using reasonable voltages and having operating lifetimes of many thousands of hours. Methods are being developed to permit economical manufacture. New techniques are being exploited to realize lightweight sources suitable for airborne and space use.

It is convenient to separate the cw sources into low or high power categories. Since there seems to be a relative abundance of sources delivering tens of milliwatts, but very few delivering over 1 watt of cw power, a division at the 1 watt level has been chosen to separate "low" power from "high" power.

All of the commercially available low power prime sources are either backward-wave oscillators or klystrons. The power available from the prominent tube lines supplied by various manfacturers is shown in Figures 5-3 and 5-4 for the millimeter portion of the spectrum. Brief comments on the various lines are included in the following discussion.

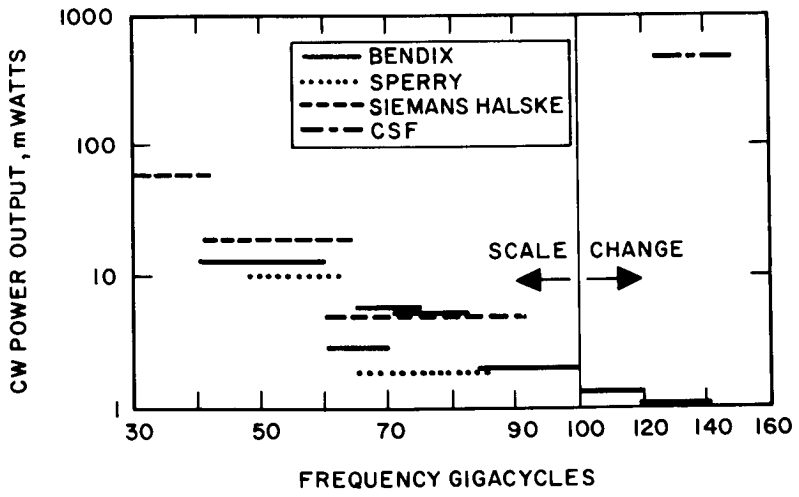


Figure 5-3. Power characteristics of available low-power backward-wave oscillators. Sperry and Bendix use Karp structure, and CSF used the vane line. In general, these tubes use low operating voltages and are suited for local oscillator and laboratory source use.

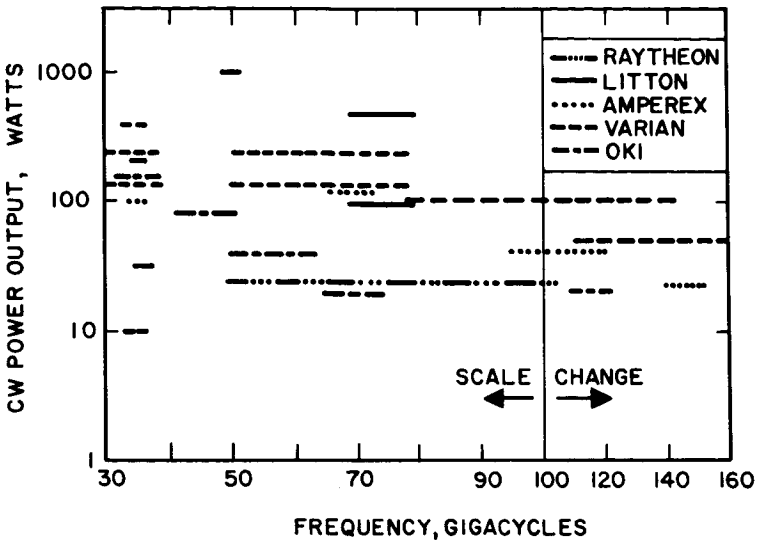


Figure 5-4. Power characteristics of available low power klystrons. All tubes are reflex klystrons and are light in weight since they use no magnetic focusing fields. They are particularly suited as pump, local oscillator, and laboratory signal sources.

Four companies market extensive lines of low power backward-wave oscillators. The most complete line is the TWO series of Bendix, which covers the entire range from 40 to 140 Gc. Most of these tubes can be procured in either solenoids or permanent magnets. They use Karp structures and can be electronically tuned over 15 percent ranges. Maximum operating voltages are 3500 volts or less.

Sperry has developed a very similar line of tubes covering the range up to about 90 Gc. Designations are SBM 421 and SBE 402. These tubes are packaged in permanent magnets and weigh only 7 pounds. Both tubes use the Karp structure and maximum operating voltage is 3,200 volts.

Siemens-Halske markets the RWO 40, 60, and 80 covering the frequencies from 26.5 to 90 Gc. Power output varies from 60 to 5 mw. The RWO 60 weighs about 17 pounds. These tubes require several variable voltages for various focusing electrodes. They use a form of interdigital slow-wave structure and convergent electron guns.

CSF of France dominates the very high frequency range beyond that shown in Figure 5-3. The COE-20 will deliver 500 mw over a 10 Gc range at about 140 Gc, and will deliver about 1 watt over a few gigacycles. This is the first tube discussed which uses the Millman structure and is more typical of the high power designs. The COE 10 delivers 10 to 20 mw over a 10 percent range at 300 Gc. The COS-09 delivers 30 to 50 mw in the 350 Gc range. The COS-07 delivers 5 to 10 mw in the 400 Gc range, while the COS-06 delivers 5 mw in the vicinity of 485 Gc. This company (CSF) has demonstrated an oscillator which delivers 1 mw at 708 Gc, the highest frequency oscillator ever generated by this means. All of these tubes use the the same general structure together with highly convergent electron guns. For the most part, operating voltage is kept to 7,000 volts or less

Many companies market extensive klystron lines operating up to 170 Gc. Most extensive coverage is achieved by Varian, which lists tubes capable of delivering 100 mw or more to 140 Gc, and 50 mw to 170 Gc. Maximum voltage is 2500 volts. The tubes are extremely light in weight and are air cooled.

Oki Electric of Japan lists a series of reflex klystrons covering the range from 30 to 100 Gc. Available performance data indicate power levels of over 100 mw in the 30 Gc range to about 60 mw in the 75 Gc range.

Amperex (Philips) lists the DX 184, 151, 242, and 237 which operate at 8 mm, 4 mm, 3.2 mm, and 2.5 mm, respectively. They deliver several tens of milliwatts up to 100 mw.

Raytheon markets an extensive line with frequency coverage to about 120 Gc in its QKK series. Power levels vary from 100 mw at the lower frequencies to 20 mw at the higher frequencies.

Litton Industries also markets several relatively high powered reflex klystrons in the 35, 50, and 70 Gc ranges.

Tubes listed below deliver 1 watt or more. All of the tubes which meet this requirement are listed in Figure 5-5 and all except CM08X are linear beam, O-type devices.

The type number of each is shown followed by the letters O or A, to designate it as an oscillator or amplifier.

All of the oscillators shown in Figure 5-5 are of the floating drift tube klystron type or of the Millman backward-wave oscillator type, or are closely related. Hughes lists backward-wave oscillators delivering 15, 8 and 2 watts at 5.5, 3.2, and 2 mm, respectively. All of these tubes are air cooled and operate at efficiencies up to 15 percent. All use depressed collectors to achieve this relatively high efficiency and low tube dissipation.

Litton markets the Elliott line of floating drift tube klystrons which operate in the 35 and 50 Gc ranges. The highest power available is 50 watts at 35 Gc from the L3736. All of the tubes are liquid cooled and mounted in permanent magnet packages weighing about 10 pounds.

Oki lists the 35F10 and 50F10 laddertrons which deliver 10 and 2 watts, respectively. Their principle of operation is quite similar to the floating drift tube klystron except that a distributed interaction is provided in the cavity. CSF lists another Millman type of oscillator which delivers 8 watts over an extensive range at 70 Gc. Watkins-Johnson advertises the W-J 282 which delivers 750 watts at 35 Gc. It is a modified version of the W-J 266 amplifier which includes an internal feedback mechanism to make it oscillate.

Hughes lists the 812H, 813H, and 814H traveling-wave amplifiers which deliver 150 and 1200 watts at 5.5 mm, and 150 watts at 3.2 mm. All of the tubes use coupled-cavity slow-wave structures and highly convergent electron guns. The 812H and 814H are air cooled while the 813H requires liquid cooling. All of the tubes use depressed collectors (operating at about 30 percent of beam potential) to maximize efficiency. The 812H and 813H operate at 30 to 35 percent efficiency, while efficiency of the 814H is 20 percent. Small signal gain of the 813H is 26 db, with saturated gain over 20 db.

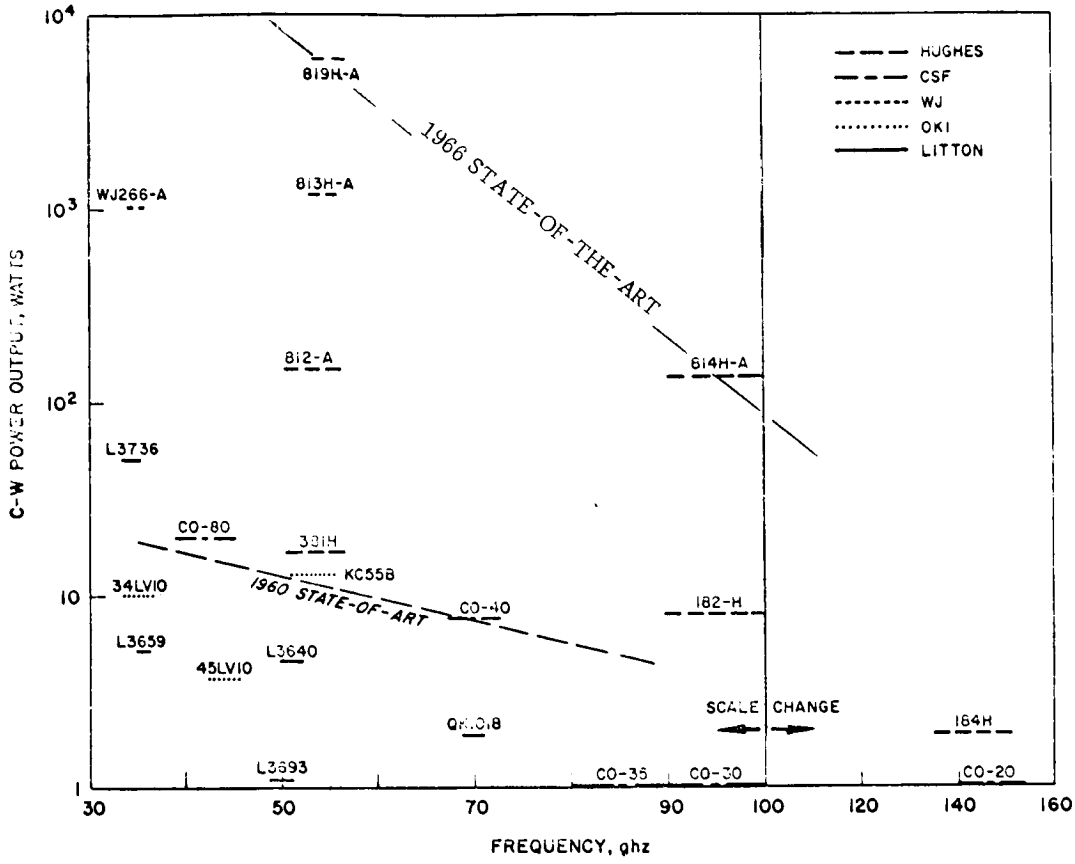


Figure 5-5. Power characteristics of available high power cw sources. Most significant is 35 percent efficiency achieved in the 813H and marked increase in available power which has occurred since 1960. The letter -A stands for amplifier and the letter -O signifies oscillator.

Millimeter Wave Sources. Two new tubes have been developed which have significantly advanced the state of the art of high power generation. The Hughes 819H has demonstrated over 5 kw of average power output at 55 Gc with a large signal gain of more than 20 db. Collector depression results in efficiencies of about 30 percent.

The Hughes 826H tube has shown output power of almost 600 watts at 94 Gc. This tube has been operated in excess of 10 percent duty cycle with such very long pulse widths that the design is believed to be fully capable of cw operation.

With reference to Figure 5-5 these results obviously describe a new boundary to the state-of-the-art. In the past six years, the available output power has been increased by approximately two orders of magnitude.

The W-J 226 also uses the coupled cavity circuit and delivers 1 kw at 35 Gc. Gain is 13 db and operating efficiency is about 10 percent. This tube is also liquid cooled but does not incorporate a depressed collector.

Solid State Sources. Most active microwave solid state devices can be used either as oscillators or amplifiers, depending upon the associated external circuitry and choice of operating point on the device characteristic. For the purpose of this discussion, only the oscillator mode of operation will be considered. The various solid state oscillator sources are listed as follows:

1. Transistor oscillators
2. Transistor oscillator - varactor multiplier combination
3. Parametric oscillators
4. Tunnel diode oscillators
5. Impact avalanche transit time oscillators
6. Gunn effect oscillators

Parametric oscillators are not of prime interest since they require auxiliary high frequency pump oscillators. Tunnel diode oscillators operate at high frequencies but appear to be power limited. Thus they will be excluded from further consideration in this discussion. The power generation capabilities of transistor oscillators and associated varactor multipliers will be summarized briefly. However, the bulk of the discussion will be devoted to descriptions of the IMPATT (IMPact Avalanche Transit Time) and Gunn oscillators which appear at this time to offer potential for higher power levels at high efficiency.

Transistor Oscillators. Extensive efforts have been underway for many years to improve the power-frequency characteristics of transistor oscillators. Recent progress has been achieved through use of paralleled transistors, monolithic integrated circuit techniques, and

incorporation of improved varactor multipliers. The current state of the art as summarized by Matther[1] is shown in Figure 5-6. The stimulus for study of the IMPATT and Gunn oscillators is provided by the fact that the power frequency performance of Figure 5-6 is already exceeded in many instances with rudimentary versions of these more recent devices.

IMPATT Oscillators. In general, these oscillators involve use of a semiconductor biased in the reverse direction and mounted in a microwave cavity. Bias is applied such that the operating point occurs in the avalanche breakdown region of the diode characteristic. Negative resistance and oscillations have been obtained in simple reverse biased p-n junctions as well as in the more complex n^+pip^+ configuration as first proposed by Read.

Gilden and Hines[2] have depicted the typical p-n junction in reverse bias for avalanche conditions as shown in Figure 5-7. The active zone is composed of the thin avalanche region followed by the depletion zone through which carriers drift at the saturation velocity V_D. The realization of a negative resistance is associated with a 90-degree phase lag between current and applied ac voltage which occurs in the avalanche process followed by a further 90-degree lag during the transit time of the carriers through the depletion zone.

Hines has shown that the equivalent impedance of the diode can be expressed as

$$Z = R_s + \frac{\ell d^2}{V_D \epsilon A} \frac{1}{\frac{1-\omega^2}{\omega a^2}} + \frac{1}{j\omega C} \frac{1}{\frac{1-\omega a^2}{\omega^2}} \qquad (5\text{-}1)$$

under the condition that the transmit time through the depletion layer (expressed in radians)

$$\theta = \frac{\omega \ell_D}{V_D} = \omega \tau_D$$

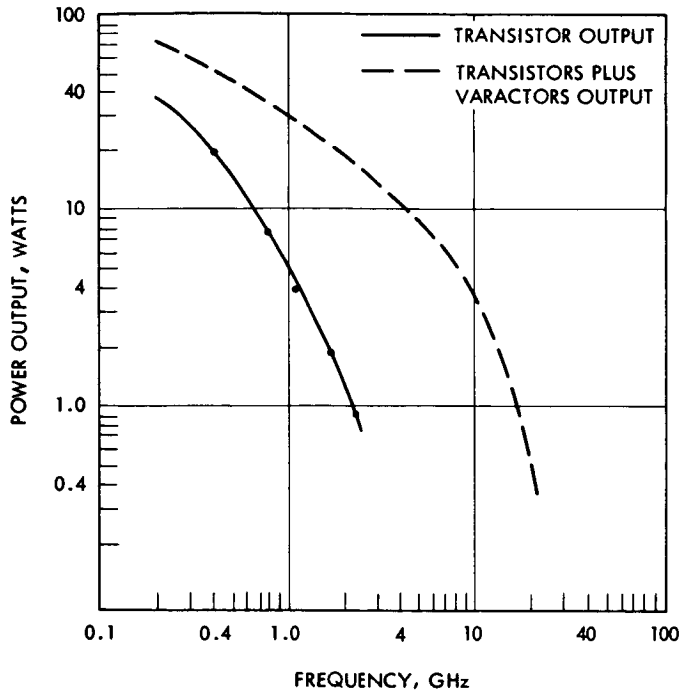


Figure 5-6. State of the art of the power output of primary transistor oscillators and transistor varactor-multipliers versus output frequency.

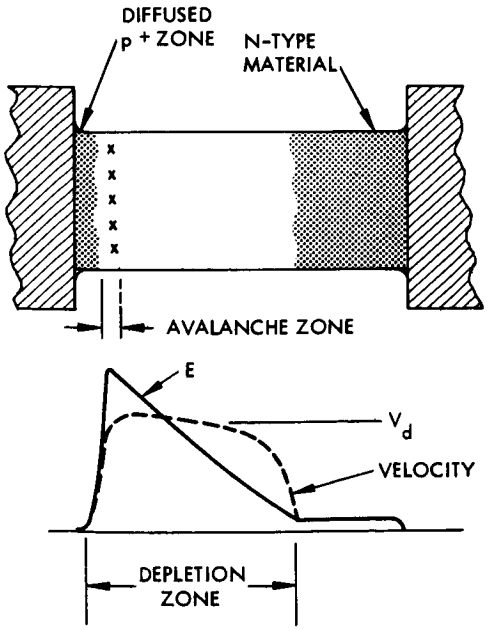


Figure 5-7. Typical P-N junction in reverse bias (conditions at avalanche).

is less than $\pi/4$. R_s is the bulk resistance of the region outside the depletion zone, ℓ_o is the length of the depletion zone, A is the junction area, and C is the capacitance of the depletion and avalanche zones, i.e.,

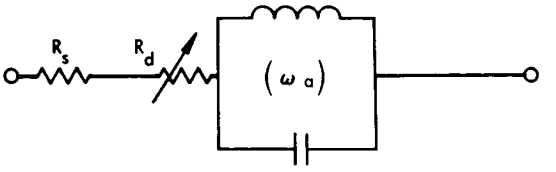
$$C = \frac{\epsilon A}{\ell_a + \ell_D}$$

ω_a is a parameter which is denoted as the avalanche frequency and is given by

$$\omega_a^2 = \frac{2\alpha' V_D I_o}{\epsilon A}$$

where α is the ionization coefficient and the prime denotes the derivative with respect to electric field, and I_o is the bias current. It is important to note that ω_a is a function only of the materials, junction geometry, and bias current. The equivalent circuit suggested by equation (5-1) is shown in Figure 5-8, where the variable resistance R_D represents the second term and the impedance of the tuned circuit the third term. Note that for $\omega < \omega_a$, R_D is positive and the tuned circuit presents an inductive reactance. For $\omega > \omega_a$, R_D becomes negative and the reactance becomes capacitive. This dependence is sketched in Figure 5-9. The oscillator is formed by supplying an inductance L in parallel through microwave circuitry; the negative of this inductive reactance is plotted in Figure 5-9. The oscillator will operate at frequency ω_r where the diode capacitive reactance is equal to the inductive reactance; ω_o is the resonant frequency of the system at zero bias current. Tuning of the oscillator is accomplished either by varying I_o (and thus ω_a) or by varying L through mechanical adjustment of the microwave cavity. The operating frequency of the IMPATT oscillator will be given approximately by

$$f \approx \frac{V_D}{2\ell_D}$$



$$C = \frac{\epsilon A}{\ell_d + \ell_a}$$

Figure 5-8. Equivalent circuit for avalanche diode at small transit angle.

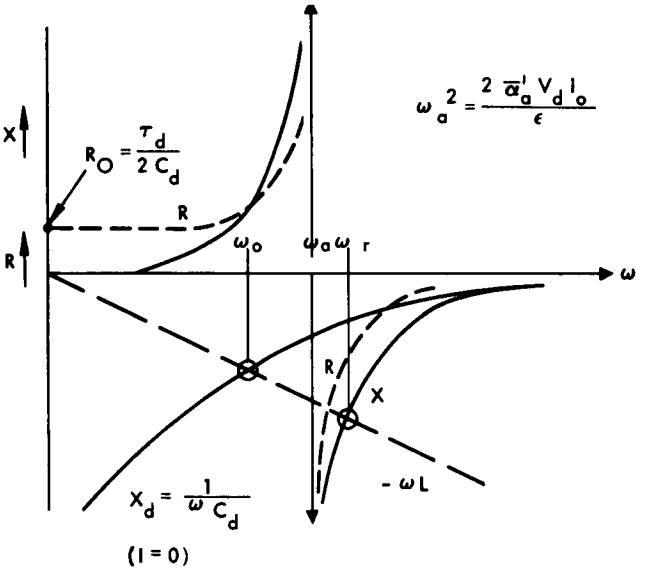


Figure 5-9. Frequency dependence of the real and imaginary parts of the diode impedance.

Since V_D is approximately twice as high for gallium arsenide as compared with silicon, the former will be more desirable for higher frequency operation. On the other hand, silicon technology is more advanced, and the higher defect density and higher thermal resistivity of gallium arsenide may limit its power capability.

The current state of the art in power generation from IMPATT oscillators is shown in Figure 5-10. Ultimate performance of these oscillators remains in doubt; many of the current results are being obtained from off-the-shelf varactor diodes which are not optimized for the application.

An analysis of the anticipated noise performance when used in amplifiers predicts noise figures of 37 to 40 db. Accordingly, it is anticipated that IMPATT oscillators will be relatively noisy. Brand[3], et al., have observed a relatively high noise level in gallium arsenide oscillators. They note a significant increase in the noise figure of a single-ended mixer using the IMPATT oscillator as a local oscillator when compared with results using a klystron. However, in a balanced mixer configuration, they are able to duplicate the performance obtained with the klystron.

In summary, the characteristics of the IMPATT oscillators are as follows:

1. Relatively high power-frequency characteristic
2. Frequency tuned by external circuit
3. Electronically tuned by bias current
4. Relatively high noise content in output

Gunn Effect Oscillators. Unlike the IMPATT oscillator, the Gunn oscillators do not require the formation of a junction. The effect takes place in a sample of the bulk material which is supplied with ohmic contacts on each side. Typical length of the samples is 25 to 200 μ. Gallium arsenide is the most common material; however, the effect has been observed in cadmium telluride and indium arsenide. Typical doping densities for gallium arsenide are 10^{13} to $10^{14}/cm^3$.

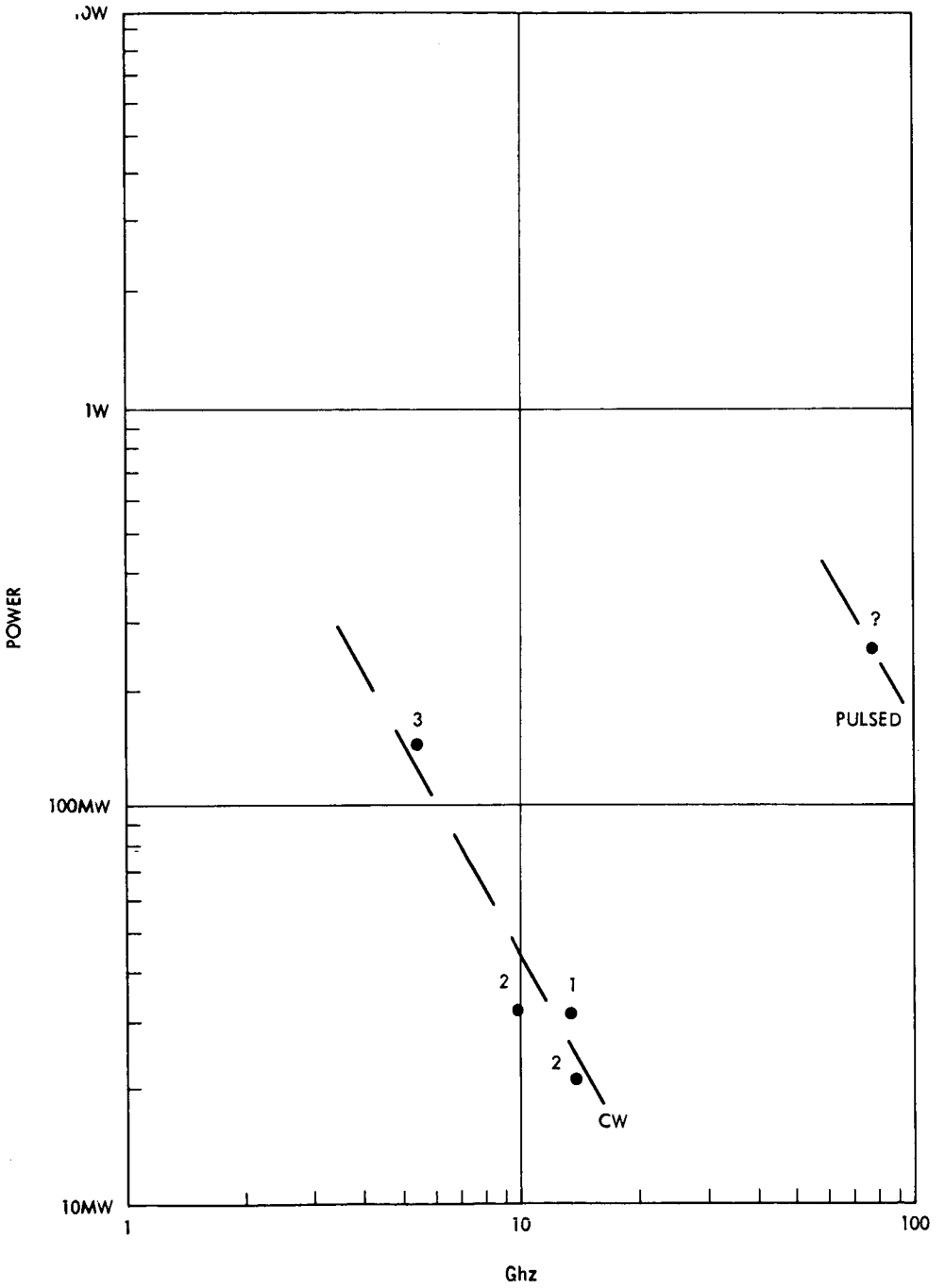


Figure 5-10. IMPATT performance.

Two modes of operation associated with the general "Gunn effect" concept have been observed. The first, as originally observed by Gunn[4], takes place in the parameter range

$$\rho L > 2 \times 10^{12} \text{ cm}^2$$

where ρ is the doping density and L the length of the sample. A second mode of operation as observed by Thim[5], et al., occurs in the range

$$2 \times 10^{12} > \rho L > 10^{10} \text{ to } 10^{11} \text{ cm}^2$$

Performance characteristics for these modes will be described later in this discussion.

The Gunn effect is essentially a current instability which occurs when an electric field of 3 to 5 kv/cm is applied across the bulk semiconductor. Associated with this field is saturation of the drift velocity of the carriers, thus removing the response out of the ohmic range. Increasing the field slightly beyond the onset of current saturation results in an instability which is manifested by large periodic spikes in the current.

This effect is explained for gallium arsenide by a decrease in average mobility of the carriers as field is increased beyond the threshold. In the energy band structure of gallium arsenide, in addition to the low lying principal conduction band, there are six satellite bands lying approximately 0.36 eV above the bottom of principal band. The mobility of electrons in the principal band is much higher than for that of the satellite bands. When the electric field reaches the threshold value, many electrons acquire sufficient energy to transfer to the satellite conduction bands where their mobility is low. Thus, on the average, a decrease in mobility occurs, and consequently a reduction in current, or, equivalently, a negative resistance characteristic evolves.

When the applied field is above threshold, then it can be shown that a high field region (domain) with two low field regions are created within the sample. These domains then travel the length of the sample in a transit time .

$$\tau = \frac{L}{V_D}$$

The domains are created at the negative electrode, travel across the sample to the anode where they disappear simultaneously with the formation of the next domain. The fundamental frequency of the modulated current which results is given by

$$f = \frac{1}{\tau} = \frac{V_D}{L}$$

Since the frequency is dependent upon transit time, it is essential to have a planar surface at the negative electrode. Curved surfaces can lead to erratic domain formation in space and thus erratic output frequency.

The preceding theory for the current instabilities explains the observed performance in the first mode of operation, i.e., in the range $\rho L > 2 \times 10^{12}$ cm. The length of the sample for this range is typically greater than 100 μ. This mode is characterized by

Spiked current waveforms
High harmonic content
Output frequency not influenced by circuit load
Relatively high output power

The second mode of operation is associated with instabilities in the range $\rho L < 2 \times 10^{12}$ cm^2, with sample lengths typically 25 to 100 μ. In this mode, the instabilities are associated with the formation of space-charge waves in the sample. Growing space-charge wave effects have been observed by several investigators in this range. Hakki[6]

et al., have observed the transition from one mode to the other in a single sample. The second mode is characterized by

Sinusoidal current waveforms

Harmonics 25 to 80 db below fundamental

Output frequency tuned by external load

Small signal range (as compared to first mode)

The state-of-the-art for power generation by Gunn oscillators is shown in Figure 5-11. These results have been obtained from oscillators operating in the first mode. Again it is difficult to anticipate ultimate performance from these relatively new devices. However, speculative estimates of pulsed powers range from 10 kw at L-band to 100 watts at X-band.

Currently, the consensus of opinion seems to favor the IMPATT oscillator as the most useful on the basis that its power frequency characteristic is higher, its frequency can be adjusted by external circuitry and bias current, and the width of a depletion zone can be made extremely small with relative ease as compared to the thickness of bulk samples. Noise properties favor the Gunn effect oscillators, however, so that both types may find extensive use.

5.2.3 Power-Burden Relationships

The evaluation of radio frequency transmitting sources was made as a function of frequency. The total amount of power per unit area was considered as a function of the cost of producing that power. For a spaceborne source, this cost involved the weight and prime power requirements. For a ground based transmitter, these costs were not considered because of the availability of space and power. The only fundamental limitation is, therefore, antenna gain which varies as the square of the frequency. This, of course, assumes no limit to the mechanical accuracy to which the antenna can be constructed nor a limit to the pointing accuracy.

The evaluation of CW radio frequency generation equipment according to the figure of merit, $M_{SB} = \frac{\eta^2}{\lambda^2} \frac{P}{W}$ and $M_G = \frac{P_{max}}{\lambda^2}$, was considered over the frequency range of 1 to 300 Gc. Using rough

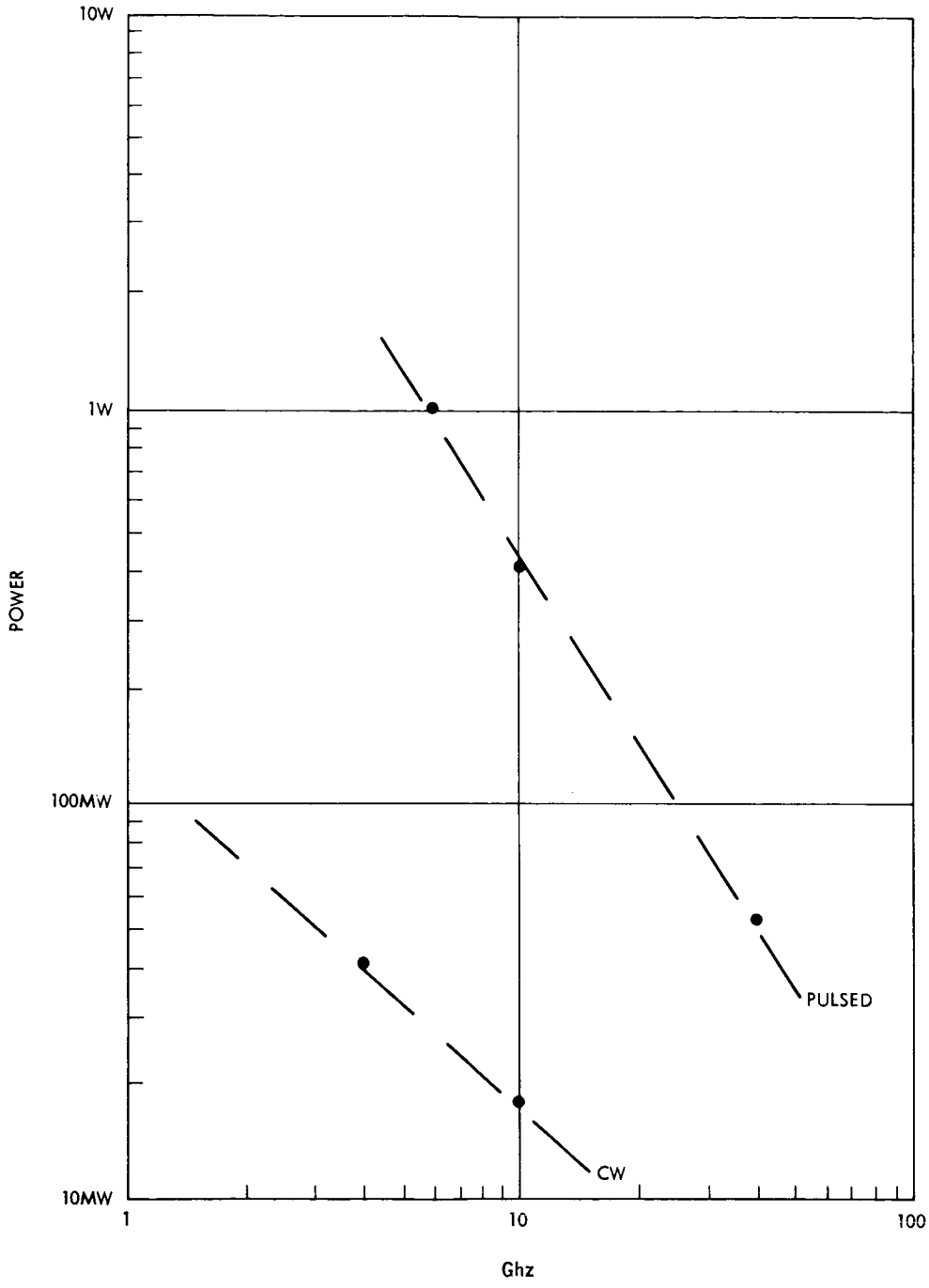


Figure 5-11. Gunn oscillator performance.

approximations of the typical numbers of the quantities in the formula, the significant trends were clearly evident. Over the specified frequency range, each parameter changed approximately as follows:

efficiency = η^2, changed about one order of magnitude.

power (watts) = P, changed about six orders of magnitude, decreasing from about 1 megawatt to 1 watt with increasing frequency.

wavelength (meters) = λ^2, changed about five orders of magnitude, decreasing from 10^{-1} to 10^{-6} meters^2 with increasing frequency.

weight (pounds) = W, changed about two orders of magnitude, decreasing from about 10^3 to 10 pounds.

It is obvious that the primary characteristics of the figure of merit are due to the available power and the antenna gain. This condition is helpful in the evaluation because these two quantities can be determined in a survey with greatest confidence.

In general, the highest figures of merit were obtained with triodes at the lowest frequencies, with klystrons in the mid-microwave area, and traveling wave tubes in the millimeter wave spectrum. The extremely light, low power tubes never competed with the heavy, brute-force type of high power device in this evaluation.

In Figure 5-12, the figure of merit M_{SB} is plotted as a function of frequency along with plots of the two separate terms included in the overall figure of merit. The dotted line takes into account some tubes which are currently under active development with enough results to justify their inclusion in this analysis.

The two portions of the figure of merit expression are plotted separately to illuminate the characteristic more fully. The expression η^2/λ^2 is basically a term over which the designer has very little control. Even major advances in the state-of-the-art in the improvement of efficiency would not alter this curve seriously. Of course, there is no control over the λ^2 term. Present design efforts are heavily engaged in improvements in power and weight so this curve is plotted separately. The composite figure of merit has an optimum value at

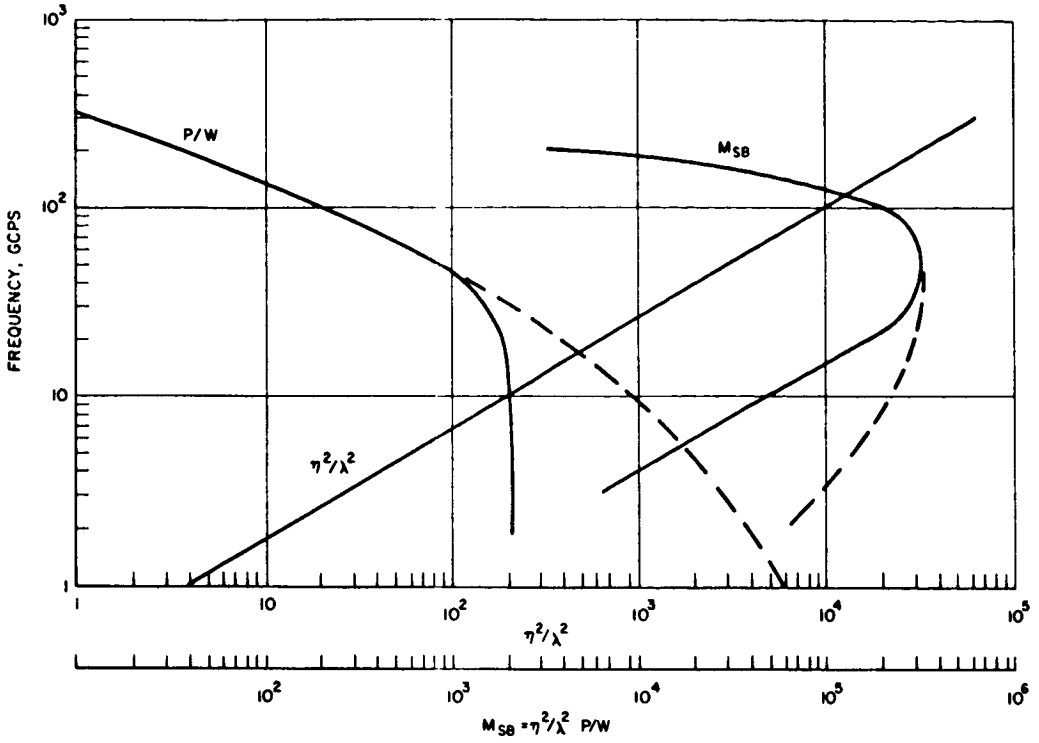


Figure 5-12. Figure of merit for spaceborne cw transmitter.

about 35 Gc which also corresponds to an excellent atmospheric window.

Figure 5-13 depicts the figure of merit M_G for a ground based transmitter as a function of frequency. Once again, the dotted line represents the ever increasing state-of-the-art as represented by current development.

This characteristic also shows an optimum value of frequency aroung 35 Gc.

5.3 OPTICAL FREQUENCY TRANSMITTING SOURCES

5.3.1 Optical Frequency Source Theory

This section presents, in brief form, the general theory of laser oscillators and amplifiers, resulting in general relationships between some operational characteristics (optical gain, gain saturation, noise)

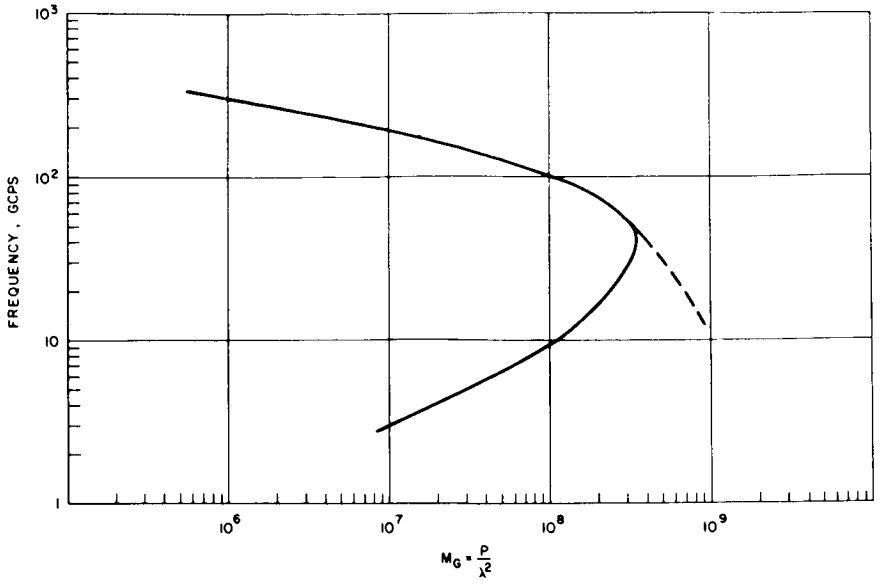


Figure 5-13. Figure of merit for ground based cw transmitter.

and some physical and atomic parameters (length, transition probabilities, linewidth, etc.). This discussion will apply, as much as possible, to all gas laser sources*.

Following this general introduction, the more detailed theory of selected gas laser sources is presented, including the considerations that are unique to that particular laser. While the selection of lasers to be treated is made on the basis of practicality, the individual discussions will usually be representative of a class of lasers. For example, the theory of the blue and green transitions in singly-ionized argon will generally apply to all singly-ionized laser transitions.

In addition, an attempt will be made to indicate the confidence with which the theory is held. For example, while the relevent processes are reasonably well known for the argon and helium-neon lasers, there are, at the present time, several conflicting theories about the role of N_2 and He in the CO_2-N_2-He laser. It is expected that these

*The theory presented in this report treats exclusively gas lasers. The theory of solid state and semi-conductor lasers will be treated in subsequent reports.

sections on the theory of individual lasers will have to be revised from time to time as the program progresses.

A simplified model of a 4-level gas laser system was used by Gordon, White and Rigden to analyze gain saturation in a 3.39 micron helium-neon laser. The treatment is valid, however, for all gas lasers in which Doppler broadening predominates.

The model treated by Gordon, et al., results in an expression for the gain of a single-pass gas laser amplifier as a function of the intensity of the signal passing through it. The expression requires knowledge of transition probabilities and other physical parameters of the lasering gas and also requires knowledge of two phenomenological "production rates" for upper and lower laser levels. These "production" terms include all the processes leading to the creation and destruction of upper and lower laser levels. The relationships between these production terms and the gas laser plasma parameters and excitation method is, necessarily, discussed under each particular laser, since, they are quite different from laser to laser. In that sense, the present model is really a "sub model." The predictions that can be made from the present model are in good qualitative agreement with experiment and thus confirm its validity as far as it goes. A brief review of this "submodel" is given below.

Consider the processes occurring in a Doppler broadened laser line centered about frequency ν_o when an incident wave of frequency ν passes through it. The rate equations (which assume a four-level

system*) for those excited atoms whose Doppler shifted center frequency is ν' can be written:

$$\begin{aligned}
\dot{n}_3(\nu', z) &= \Sigma_3(\nu') - n_3(\nu', z)\left[A_3 + B'_{32}(\nu', \nu)\, S(\nu, z)/4\pi\right] \\
&\quad + n_2(\nu', z)\, B'_{23}(\nu', \nu)\, S(\nu, z)/4\pi \\
\dot{n}_2(\nu', z) &= \Sigma_2(\nu') + n_3(\nu', z)\left[A_{32} + B'_{32}(\nu', \nu)\, S(\nu, z)/4\pi\right] \\
&\quad - n_2(\nu', z)\left[A_2 + B'_{23}(\nu', \nu)\, S(\nu, z)/4\pi\right]
\end{aligned} \tag{5-2}$$

The quantities n_3 and n_2 are the volume densities of atoms in the upper and lower laser levels that will interact with radiation at frequency ν'. $S(\nu, z)$ is the intensity of the signal at a position z from the input of the laser tube. Σ_3 and Σ_2 are phenomenologically invented pumping terms, giving the rate of production of atoms in upper and lower levels. A_3 and A_2 are the net total spontaneous emission coefficients for upper and lower laser levels. A_{32} is the spontaneous transition rate from upper to lower level. The coefficients $B'_{32}(\nu', \nu)$ and $B'_{23}(\nu', \nu)$ are the stimulated emission rates for atoms whose Doppler shift is ν' when the signal frequency is ν.

The growth rate of the signal as it is amplified by the laser can be written

$$\begin{aligned}
\frac{dS(\nu, z)}{dz} = \frac{h\nu S(\nu, z)}{4\pi} \int_0^\infty \Big[& B'_{32}(\nu', \nu)\, n_3(\nu', z) \\
& - B'_{23}(\nu', \nu)\, n_2(\nu', z)\Big]\, d\nu'
\end{aligned} \tag{5-3}$$

*The requirement of a four-(or more) level system is implicit in these equations, even though they are written for only two of the four levels. Any attempt to apply these to a three-level system would introduce a constraint on A_2 and A_3 which does not occur in all the gas laser systems considered.

for steady state (cw) operation $\dot{n}_3 = 0$ and $\dot{n}_2 = 0$, so that (5-2) and (5-3) may be combined to give a differential equation in $S(\nu, z)$. The next step is to integrate over all possible Doppler shifts $\nu_o - \nu'$ and over the length of the tube z, to find the over-all gain of the laser as a function of the frequency and intensity of the incident radiation.

5.3.2 Theory of the Argon Ion Laser

Excitation Processes. The model which seems to fit best the observed behavior of the argon ion laser is a three-level system originally proposed by Gordon, Labuda, and Miller. In this model, the upper laser level is populated by two successive electron collisions; the first produces an unexcited ion from a neutral atom, and the second excites the ion to the upper laser level. This two step process is consistent with the observed I^2 dependence of spontaneous emission from the singly ionized laser upper levels and the I^4 dependence of power output. (The reasons for the I^4 dependence are discussed in the Scaling Laws paragraph for Ion Lasers.) The depopulation of the lower laser level then occurs by vacuum ultraviolet radiation to the ion ground state; this suggests a lifetime for the lower level roughly $(\lambda_{\text{vac UV}}/\lambda_{\text{laser}})^3$ shorter than the upper level.

The population processes are shown schematically in Figure 5-14. It is not known directly whether the atom makes a "round trip" to the neutral ground state for each laser photon emitted. (This is an important factor in the laser's efficiency, since the electron energy that goes into the creation of the ion is essentially wasted.) However, the observed I^2 and I^4 dependences seem to indicate that it does. There is also some evidence that ionic metastable levels may play a role in the second electron collision, so that the picture of the three level system shown may be oversimplified.

The ultraviolet radiation which depopulates the lower level can be increasingly trapped as the ion density builds up. The "dead" region which develops at higher currents is a region of high attenuation rather than gain and is caused by radiation trapping. The gain is quite

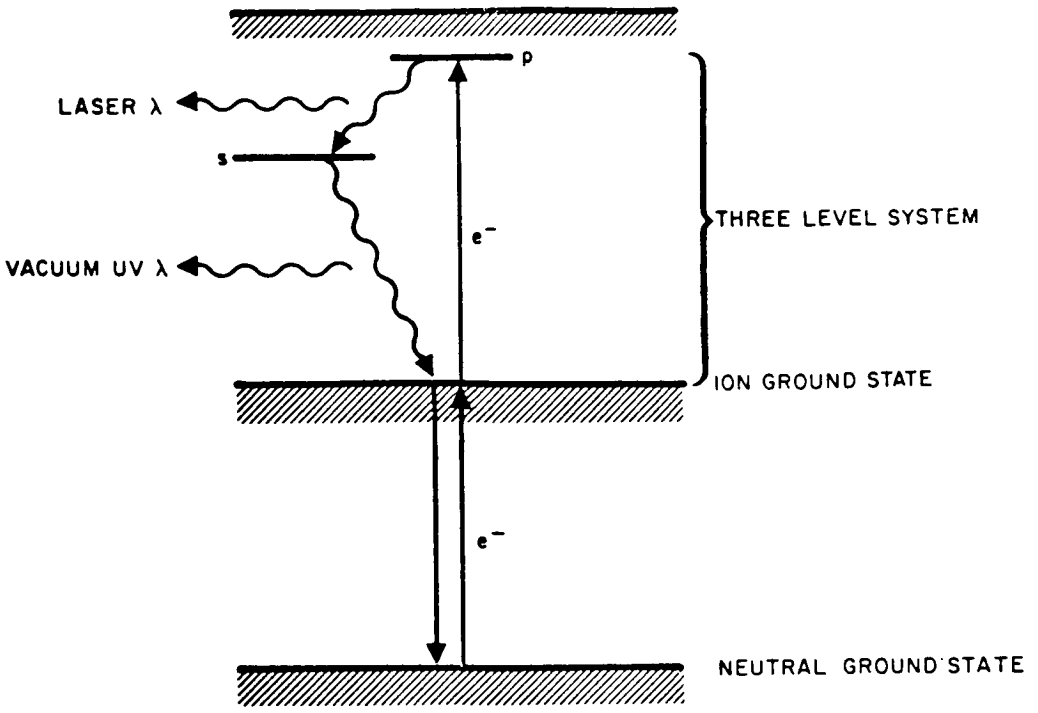


Figure 5-14. Schematic energy level diagram and processes for singly ionized atoms.

sensitive to the trapping coefficient* which, in turn, depends on the gas temperature. As the current pulse persists, the gas temperature rises and the trapping decreases, producing an inversion and oscillation for the remainder of the pulse, even if it extends to continuous operation. For lower currents or shorter pulse lengths this effect is not observed and the laser pulse follows the current pulse exactly.

Bennett, et al., have proposed that direct electron excitation from the neutral atom ground state to the upper laser level is the dominant populating mechanism. The preponderance of laser lines with p upper states follows from the selection rules of the "sudden perturbation" process referred to by Bennett. The short radiative lifetimes of the s and d lower levels then guarantee an inversion. From the evidence

*Briefly, when trapping is included in the rate equations, the current-independent part of the expression for the population inversion appears as the small difference of two large terms, one of which contains the gas temperature. Changes in the gas temperature can then swing the population difference from attenuation to amplification.

available at the present time, it seems likely that this population mechanism is dominant only in pulsed discharges with high E/p and short pulse duration; these are not conditions conducive to high efficiency or high average power. If very short pulse lasers (nanoseconds) are desired, then this mode of excitation may prove interesting.

In general, we may say that the excitation and de-excitation mechanisms are reasonably well identified for cw argon ion lasers. However, quantitative measurements of the relative importance of the various processes and construction of a numerically accurate mathematical model will require further detailed experimental study.

Magneto-Optical Effects in the Ion Laser. Ion lasers exhibit higher power output and higher efficiency when a longitudinal magnetic field is applied to the discharge to compress the plasma and reduce electron loss to the walls. However, the use of greater magnetic fields for even stronger plasma confinement cannot improve the performance of the laser indefinitely. In fact, there appears to be a definite optimum for an applied magnetic field, representing a compromise between the improvement gained through discharge confinement and the deleterious effects of the atomic processes which occur for higher magnetic fields. This section treats that which is known about the atomic effects of the magnetic field.

The externally applied magnetic field splits both the upper and lower laser levels which correspond to the different orientations assumed by the total spin of the atom with respect to the applied field. The resulting macroscopic effects of interest here may be classified as Zeeman line splitting and Faraday polarization rotation; they are intimately related by the atomic changes which occur.

The simple Zeeman effect may be thought of as splitting of the gain line into two similar Doppler-broadened lines with center frequencies separated by an amount proportional to the magnetic field. Each line is sensitive to only one circularly polarized component of light; the population inversion is effectively split into two equal parts which respond to different circular polarizations. Figure 5-15 is a sketch of

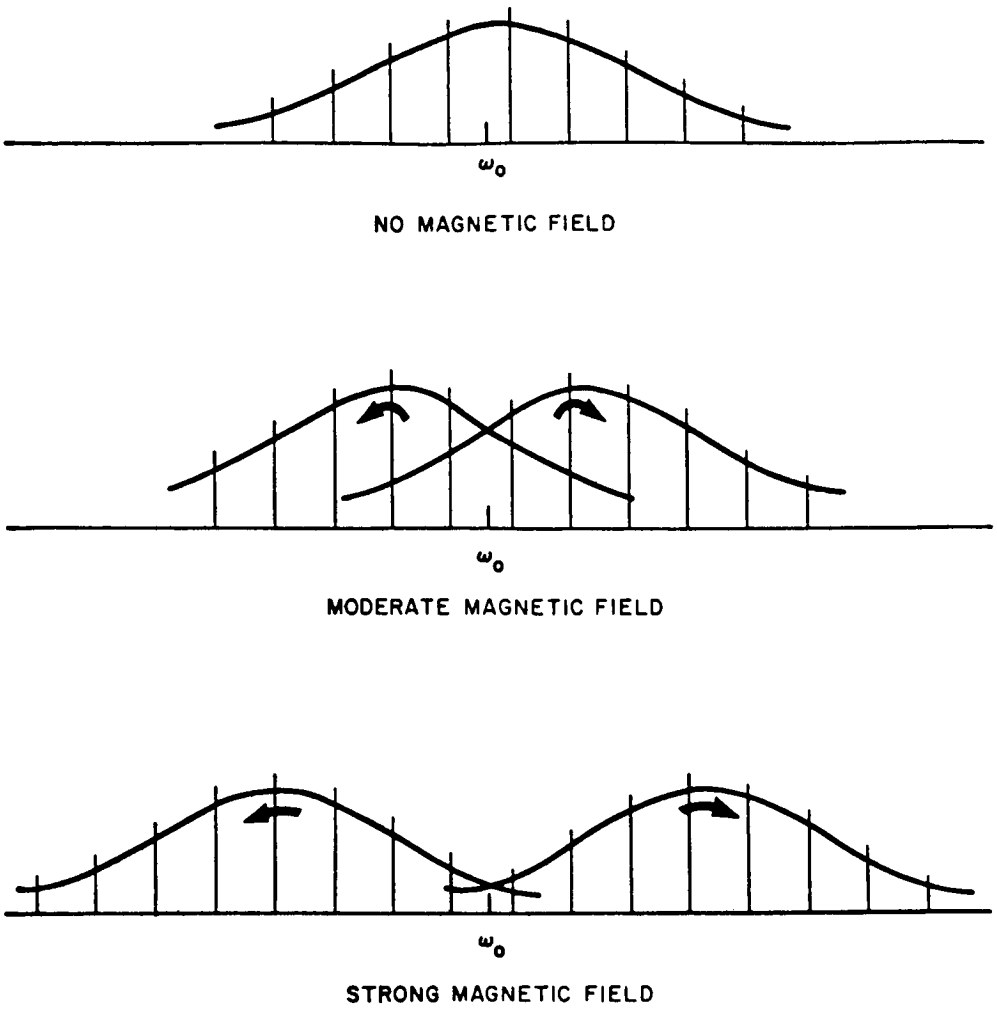


Figure 5-15. Gain curves for right and left hand circular polarization as a function of magnetic field strength.

the optical gain as a function of the light frequency at various values of applied field.

We note here that the gain which would be observed on sending a circularly polarized signal through a given length of discharge tube is not twice as great when there is no magnetic field; the maximum gain for this signal is precisely the same, but the position of the gain curve shifts in frequency as a function of the magnetic field. Thus if a laser is constructed in which both circularly polarized modes can establish themselves freely (that is, with no polarization-sensitive elements, such as Brewster windows, in the optical path), the total output power

will be largely unaffected by the line splitting, but will occur over a wider frequency range. The differential phase lag in the two polarizations, which is called Faraday rotation in the plane-polarized mode description, similarly will not affect the total output power. At zero magnetic field the output power will be just that obtained if only a single linearly polarized mode were permitting to oscillate.

However, most gas lasers have polarizing elements in the optical cavity in the form of optical windows set at the Brewster angle. These windows guarantee plane-polarized operation, which is usually desired or required by the application. In such lasers the magneto-optical effects may become very important; these effects in themselves are sufficient to cause the saturation of power with magnetic field, even if no saturation occurs in the plasma confinement.

Let us consider a laser cavity with a mode spacing that is larger than the Doppler line width (i.e., a "single frequency" laser). When a strong magnetic field is present and the Doppler-broadened line is completely split, only a single circularly polarized component of light at a given frequency can be amplified. If this circularly polarized light then passes through an optical element with transmissions 1 and a for either component of linearly polarized light, a fraction of the same circular polarization $(1/4)(1 + a^{1/2})^2$ is transmitted. The element converts $(1/4)(1 - a^{1/2})^2$ of the light to light of the opposite circular polarization, and the remainder is lost by reflection. For the typical Brewster angle windows used in gas lasers, $a = 0.85$; a negligible amount of light is converted to the opposite circular polarization (0.2 percent), but about 8 percent of the light is lost at every surface by reflection. It is thus clear that in this extreme case of very high magnetic field, two Brewster angle windows in the optical cavity introduce an additional loss of 8 percent at each of four surfaces traversed in a single pass. This additional 30 percent loss is responsible for the observed saturation in the high field case, and normal end windows would clearly be desirable if output power were the primary consideration. Note that if windows only slightly tilted from the normal were used instead of at Brewster's angle, the output is still almost completely circularly polarized. In

fact, the laser tends so strongly to the circularly polarized mode that the output is still 80 percent circularly polarized when a is as small as 0.16.

This extreme case demonstrates the great amounts of loss introduced into the cavity by any sort of partially polarizing optical element when we use magnetic fields high enough to resolve the Doppler-broadened line completely. This complete conversion to circular polarization is not observed in practice because the fields used are not sufficiently high to resolve the line in this fashion. At fields of the order of 1000 G, the laser is still largely plane-polarized and the magneto-optical effect is better described in terms of the Faraday rotation.

For magnetic fields in this intermediate range we must consider the relative indices of refraction for right and left circular polarizations. Figure 5-16 shows these indices as a function of the optical frequcney and magnetic field.

When the line splitting is small, the gains in either circular component are approximately equal, but the difference in indices of refraction causes one circular component to be slightly shifted in phase relative to the other during a single passage through the active medium. If the light is linearly polarized and the gains in either circular component are in fact equal, the only effect is a slight rotation in the plane of polarization which is proportional to the applied magnetic field. Figure 5-16 shows that for moderate fields and centrally positioned modes the right-hand component always has a larger index of refraction than the left-hand component; thus all these stronger, central modes

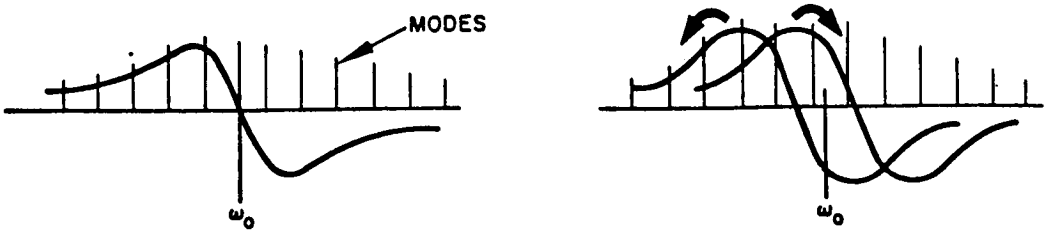


Figure 5-16. Index of refraction for right and left hand components with and without magnetic field.

exhibit polarization rotation in the same sense upon a single pass through the tube.

The rotation of polarization on a single pass through the active medium is, of course, quite small; however, the rotation adds in the same direction while the light is repeatedly reflected back through the cavity in the opposite direction. It is this multiplication of the rotation that leads to appreciable losses when polarizing elements are present in the cavity. From this effect, laboratory tubes have exhibited as much as 25 percent of the power reflected off the windows as that produced at the output end of the tube.

The actual situation is, of course, considerably more complicated than the simple picture given above. In practice, the laser tends to adopt a complicated mode which minimizes its losses as much as possible. The output from either end of the laser remains plane polarized, but is shifted from the polarization orientation with no magnetic field. This polarization shift is typically 5 to 6 percent in angle and leads to losses of some significance in gain and output power. When the splitting gets larger (approaching the gain line width), the gain on the two circularly polarized modes at any frequency becomes sufficiently different that Faraday rotation no longer gives a meaningful description of the operation, and the grosser polarization losses described previously dominate. A proper, complete description would show how these two regimes blend into one another. The over-all effect upon the loss in the cavity (and hence upon the output power of the cw laser), however, is observed to be a slowly increasing loss (by reflection at the windows) which should approach a constant at high magnetic fields.

The gain of a laser amplifier may be expressed as

$$G(S) = 1 + \frac{g_o L}{(S/S_{3dB})^n} \qquad (5\text{-}4)$$

where G(S) is the actual gain in a single pass of a signal of intensity $S(w/m^2)$ through a laser of small signal gain coefficient g_o (m^{-1}) and length L. The power level S_{3dB} is the signal power density at which the

small signal gain has been reduced by 3 dB (in the case n = 1), as shown in Figure 5-17. The exponent n equals 1 for homogeneous interaction (that is, interaction in which the signal may, with equal probability, interact with any atom in the laser), and is equal to 1/2 for inhomogeneous interaction (that is, the interaction may take place between the wave and only a small fraction of the excited atoms).

Equation (5-4) is valid if $S \gg S_{3dB}$ and if G(S) is not too large (i.e., if the approximation

$$\exp\left[\, G(S) \,\right] \approx 1 + G(S) \tag{5-5}$$

is valid). Note this does <u>not</u> require that $g_o L$ be close to unity.

Low power gas lasers usually fall into the category of inhomogeneous interactions; since the Doppler line width is so much greater than the natural line width, radiation at a single frequency can interact only with a small fraction of the atoms, e.g., those whose Doppler-shifted frequency falls within one natural line width of the incident radiation. We say this radiation "burns a hole" in the Doppler line when it depletes the excited atoms available to it without affecting the remainder.

Figure 5-18 contrasts the cases of (a) a hole burned by inhomogeneous interaction and (b) the depletion of the entire line by true homogeneous interaction. This description in terms of inhomogeneous interaction remains valid for the gas laser even when many frequencies are present (multimode operation), provided the holes do not "overlap" sufficiently (Figure 5-19). With sufficient overlapping, the gain line is effectively "burned off" (Figure 5-20) and the laser behaves as if the

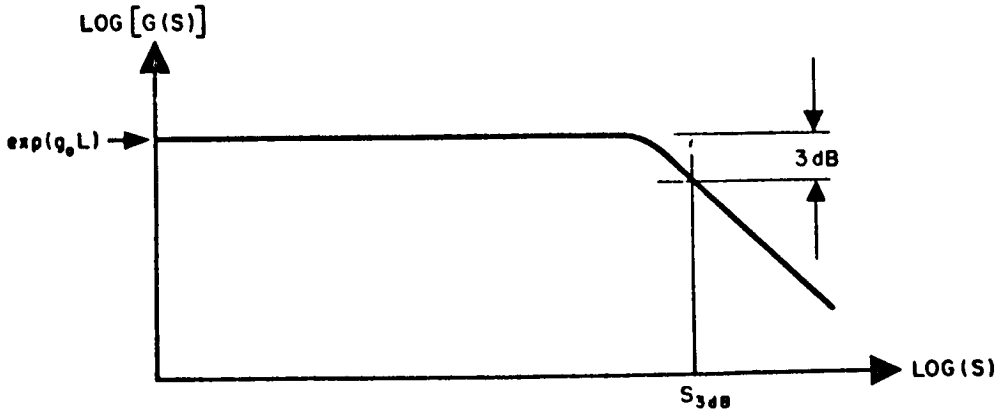


Figure 5-17. Typical gain saturation characteristic for a laser amplifier.

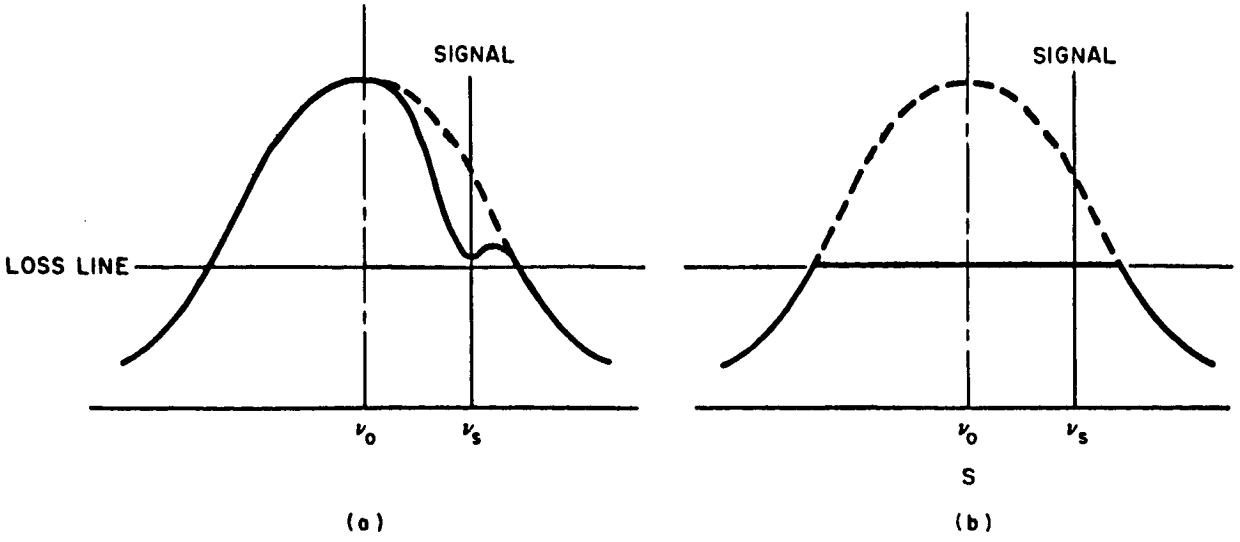


Figure 5-18. Comparison of (a) inhomogeneous interaction (hole burning) and (b) homogeneous interaction with a single frequency.

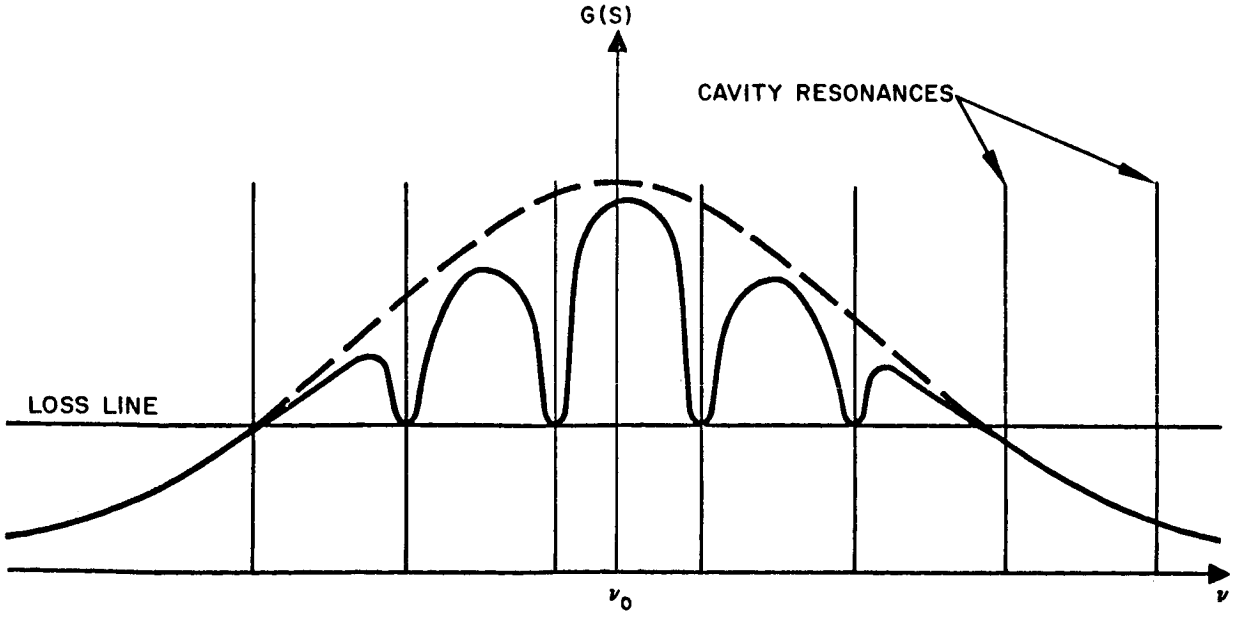


Figure 5-19. Nonoverlapping holes burned in the gain line by a multimode laser.

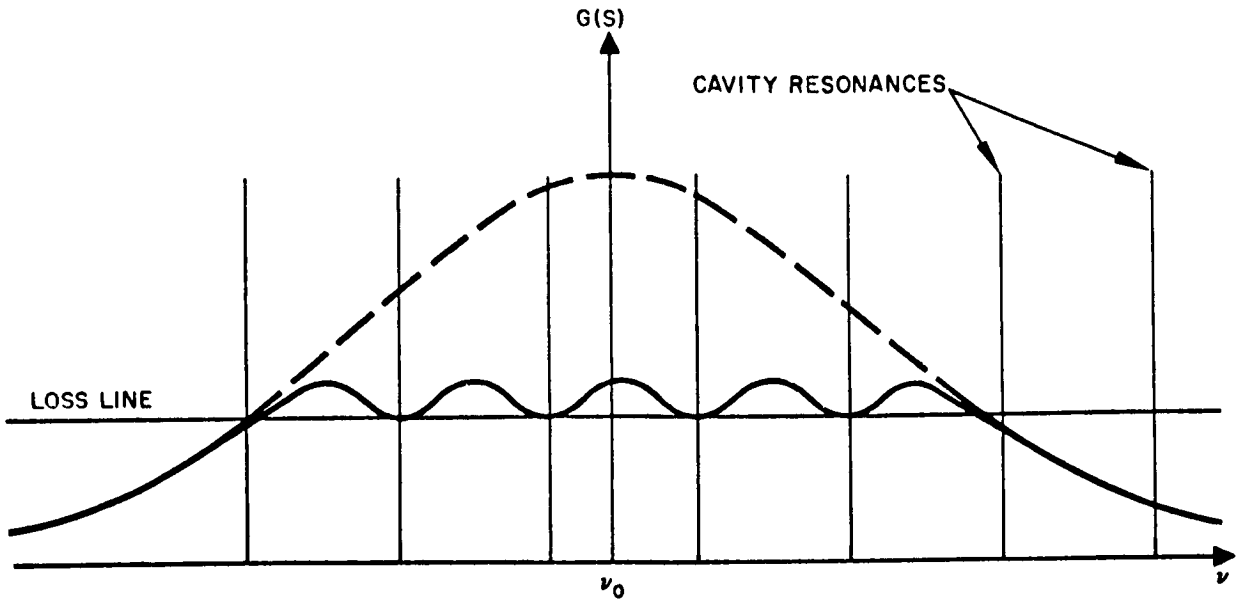


Figure 5-20. Gain line "burned off" by a multimode laser.

interaction were homogeneous. Essentially, every atom can interact with some radiation in the closely spaced multimode case*. Even if the modes are not closely spaced, the line will be burned off and behave as if the interaction were homogeneous, provided the signal intensity becomes sufficiently high. This is a result of the Lorentzian shape of the hole: it has broad "wings" which do not fall off rapidly from the hole center (not nearly so fast as the gaussian, for example). Thus a gas laser capable of high power operation under multimode conditions will exhibit inhomogeneous interaction at low signal levels and homogeneous interaction at high signal levels.

We may now use equation (5-2) to give us the variation of output power with some of the important parameters of the laser discharge. Equation (5-2) gives the gain in a single pass through the laser; if the laser is used as an oscillator, only a fraction $R < 1$ will be reflected through the tube for the next pass. If we make the simplifying

*Note that this is a physically different effect from the case of a homogeneously broadened line that typically occurs in solid state lasers (ruby, for example). In the case of a homogeneously broadened line, every atom can interact with a single frequency signal. The net effect on the gain saturation is the same, however.

assumption that losses, either dissipative or useful output, are symmetrical, the condition for steady state oscillation is

$$R\,G(S) = 1 \tag{5-6}$$

Solving (5-4) and (5-6) yields

$$S = S_{3dB}\left(\frac{Rg_oL}{1 - R}\right)^{1/n} \tag{5-7}$$

If we can further assume that $Rg_oL/(1 - R)$ is $\gg 1$ (which will be true for the important ion laser lines), equation (5-7) simplifies to

$$S \sim \begin{cases} g_oL & \text{homogeneous interaction} \\ g_o^2L^2 & \text{inhomogeneous interaction} \end{cases} \tag{5-8}$$

S is, of course, the power circulating in the cavity. The output power will be S(1 - R - losses). However, we are not concerned here with the variation with R or losses, and we assume these quantities are constant.

To relate equation (5-8) to the discharge conditions, it is necessary to know the dependence of gain on the current. The evidence for a two-step electron collision process as the key population mechanism of the upper laser level has been discussed earlier. Spontaneous emission measurements show that the number density in both upper and lower levels N_2 and $N_1(m^{-3})$ vary as J^2 (J is the current density in amperes per square meter)

$$N_2 - N_1 \sim J^2 \tag{5-9}$$

The "constant" of proportionality actually contains some implicit current dependence (viz., the effect of gas heating on the Doppler line width and the effects of radiation trapping on N_1 caused by the current dependence of the ion density N_o). However, if this variation may be neglected for the moment, equation (5-9) gives the major variation of inversion with current. The gain coefficient of a laser discharge is approximately proportional to the inversion

$$g_o \sim N_2 - N_1 \tag{5-10}$$

except very near threshold, where it varies more as the 3/2 power

$$g_o \sim (N_2 - N_1)^{3/2} \tag{5-11}$$

(Actually, this is very approximate; the actual dependence is not expressable as a simple exponent.) For a discharge tube of diameter D, we have the power output (watts) and current (amperes) given by

$$P = \frac{\pi D^2 S}{4} \tag{5-12}$$

$$I = \frac{\pi D^2 J}{4} \tag{5-13}$$

The above equations assume uniformity of S and J in the radial direction. This is probably a good assumption for J but not for S; however, to first order only the numerical coefficient (the cavity mode filling factor) will change and $P \sim D^2 S$ for a given cavity. Combining (5-8) through (5-13) we have

$$P \sim \begin{cases} D^2 g_o L \sim D^2 (N_2 - N_1) L \sim D^2 J^2 L \sim \dfrac{I^2 L}{D^2} \text{ saturated} \\ D^2 g_o^2 L^2 \sim \begin{cases} D^2 (N_2 - N_1)^2 L^2 \sim D^2 J^4 L^2 \sim \dfrac{I^4 L^2}{D^6} \text{ moderate levels} \\ D^2 (N_2 - N_1)^3 L^2 \sim D^2 J^6 L^2 \sim \dfrac{I^6 L^2}{D^{10}} \text{ near threshold} \end{cases} \end{cases} \tag{5-14}$$

The existence of $I^6 \rightarrow I^4 \rightarrow I^2$ behavior has been well verified in this laboratory; however, the exact <u>limits</u> of each regime are not well known as a function of the other parameters of the laser discharge (that is, those parameters which make up the constant of proportionality). Figure 5-21 indicates, in a highly schematic way, the expected variation

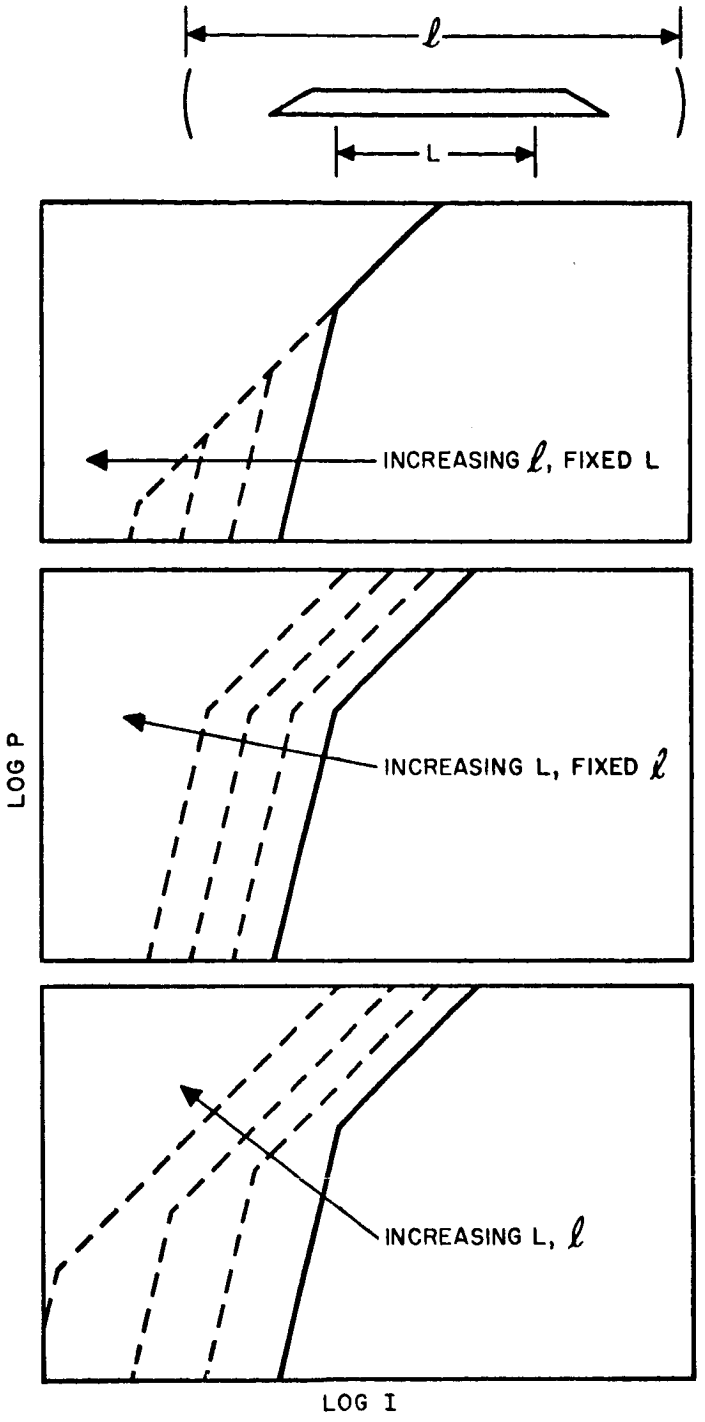


Figure 5-21. Schematic representation of the change in power with different discharge and cavity lengths.

of power output with current as the discharge length L and the cavity length ℓ are varied. The assumptions made are, of course, that $c/2\ell > \Delta\nu_N$ and $\ell > L$.

When a magnetic field is used to confine the plasma, the scaling is slightly different. Since the rate at which ions leave the discharge by diffusion to the walls is reduced by the magnetic field, the rate at which plasma is generated must also decrease in order to maintain an equilibrium state. The plasma generation rate is proportional to the electron density and the average electron energy, and the new equilibrium is maintained by a decrease in the average electron energy. The longitudinal electric field is roughly proportional to the electron energy, and it is the decrease in the longitudinal electric field (and the corresponding decrease in tube voltage) which accounts for the improvement in the efficiency of the discharge.

However, since the excitation of ions to the upper laser level is by electron-ion impact, the excitation rate is proportional to the plasma density squared and the average electron energy. We expect that while laser performance should be enhanced in a magnetic field because of the increased plasma density and efficiency of the discharge, this improvement should eventually be balanced or reversed by the decrease in the average energy of the plasma electrons.

Magneto-optical effects can further complicate the dependence of laser performance on the magnetic field, as discussed previously.

It is quite clear that one of the major first steps toward improving the power output and efficiency of the ion laser is to make a thorough and careful set of measurements of gain, gain saturation, power output, and efficiency with the physical parameters I, L, ℓ, D, B, and pressure varied over as wide a range as is practicable. Such measurements will confirm the theory of operation discussed here, explore the limits of validity of the theory, and provide a set of design criteria for future laser development.

We may also estimate the variation of efficiency with the major parameters. The power input to the discharge is approximately

$$P_{in} = IEL + IV_{cath} \tag{5-15}$$

where E is the average longitudinal electric field in the positive column and is a function of the diameter and the magnetic field. V_{cath} is the cathode fall, a function only of the cathode type and material. (For hot cathodes, such as those used in our cw tubes, V_{cath}=20 to 40 V; for cold cathodes, V_{cath} may be several thousand volts.) The exact dependence of E on diameter and magnetic field is not known. For the helium-neon laser the scaling relation pD = constant results in the relation ED = constant. Our observations to date indicate that in the Ar II laser (and in the neutral xenon 3.5 μ laser) the variation of E is faster than 1/D. The efficiencies corresponding to the three regions of operation may then be written

$$\eta \sim \begin{cases} \dfrac{I}{ED^2 + \dfrac{V_{cath}\, D^2}{L}} & \text{saturated} \\ \dfrac{I^3 L}{ED^6 + \dfrac{V_{cath}\, D^6}{L}} & \text{moderate level} \\ \dfrac{I^5 L}{ED^{10} + \dfrac{V_{cath}\, D^{10}}{L}} & \text{near threshold} \end{cases} \tag{5-16}$$

The saturated region will give the region of highest efficiency. For tubes long enough that V_{cath}/L is $\ll$ E, the saturated efficiency is independent of length, V_{cath} is typically 20 to 40 V, and E ranges from 0.5 to 20 V/cm, depending on the diameter and the magnetic field. With these numbers, it is apparent that the cathode fall will significantly degrade the efficiency in short tubes (20 cm or so).

5.3.3 Characteristics of Mode-Coupled Gas Lasers

Introduction. The atomic populations of most gas lasers are sufficiently inhomogeneously broadened to allow the simultaneous oscillation of a large number of axial modes. Each of these modes is driven by spontaneous emission from atoms in different regions of the atomic florescence line; to a first approximation, these modes are uncoupled and oscillate independently. Under normal conditions, and especially for closely spaced modes (long optical cavities), the amplitudes and phases of the individual modes will fluctuate in a random manner. The output of a multimode laser is therefore amplitude modulated and has a frequency coherence much less than that of any single axial mode.

By introducing an optical modulator within the laser cavity, the previously uncoupled modes may be coupled together so that their relative amplitudes and phases are constant in time. In gas lasers, suitable modulators vary either the optical path length of the cavity (phase modulation), or the cavity losses (loss modulation). The frequency spectrum and the time domain output of a mode-coupled laser vary with the type, frequency, and amplitude of the modulation.

The effect of mode-coupling within the laser cavity is to change the form of the laser signal. The laser output may be considered as an optical carrier, and by mode coupling the carrier is converted to a possibly more useful form, such as an unmodulated pulse train or a single frequency. These internal modulation techniques should not be confused with schemes which use a modulator within the cavity to impose information on a direct or scattered beam, such as the intracavity SSBSC modulator discussed in Section 7.0.

A general discussion of the characteristics of loss and phase modulated, mode-coupled lasers is given in the paragraphs below.

Internal Loss Modulation of Multimode Lasers. The theory of internal loss modulation has been presented by several authors[7, 8, 9], and successful mode-locked operation has been reported for He-Ne at 6328 Å[9, 10] and for A II at 4, 880 Å[9]. To introduce a time-varying loss, a suitable modulator is placed within the optical cavity as near as possible to an end mirror. The coupling effect is independent of the manner in which the losses arise, so that loss modulation may be produced by an electro-optic intensity modulator or by an elasto-optic (acoustic) modulator. With the latter type of modulator, the standing acoustic wave acts as an acoustic diffraction grating which attenuates the zero-order direct beam by scattering energy into higher order off-axis beams. Since there is a null in the standing wave twice during each period of the applied signal, the modulation of the zero order beam occurs at twice the applied acoustic frequency.

When the frequency of the time-varying loss ν_m is equal to or very close to the mode spacing c/2L, coupling of the modes results from the nonlinear polarization of the medium. The relative amplitudes of the modes becomes approximately Gaussian, centered about the line center, and all modes oscillate with equal phase. For a mode-coupled laser with a spontaneous emission linewidth $\Delta\nu$ and n axial modes oscillating in phase (normally $n \sim \Delta\nu / (c/2L)$) the laser output (as predicted by simple Fourier analysis) is a pulse train of ν_m pps with the following characteristics:

1. The pulse width is roughly $(n\,\nu_m)^{-1} \sim (\Delta\nu)^{-1}$.
2. The average power is approximately equal to the free-running laser power.
3. The peak pulse power is n times the average power.

The effect of the internal loss modulation on a multimode laser is then to convert the output to a pulse modulated signal. Experimental results [9] obtained for two types of lasers with $\nu_m = c/2L \sim 100$MHz are:

Laser	λ	$\Delta\nu$	Pulse Length	Peak Power/ Average Power
He-Ne	6328 Å	1.5 GHz	0.5 nsec	17
A II	4800 Å	4.5 GHz	0.25 nsec	20-30

Such a high-intensity pulse modulated source may have applications in PCM systems where an external optical shutter is provided to modulate the pulse height.

Internal Phase Modulation of Multimode Lasers. A mode coupling of a gas laser may also be accomplished by placing a phase modulator (KDP crystals have been used to date) inside the laser cavity. When an rf field is applied to the electro-optic crystal, its index of refraction is changed, causing a change in the length of the optical path of the cavity. This approach is equivalent to vibrating one end of the cavity mirrors at the modulation frequency, and gives rise to the generation of fm sidebands for each oscillating mode.

Two different types of mode-coupling can result depending on the frequency of the modulation. When the crystal is driven at a frequency which very nearly, but not exactly, corresponds to the frequency separation of the free running modes, the laser operates in the fm region. If the driving frequency is tuned to exactly the frequency separation of the modes, then "phase locked" operation occurs which has different characteristics from the fm region. These two modes of operation will be discussed separately below. A detailed discussion of the limits of operation of the two regions as predicted by theory [11, 12] and observed in experiments [13] is contained in the literature.

Mode Coupling in the FM region. Since the modulation frequency is roughly the frequency separation of the free running modes, the fm sidebands of each mode very nearly coincide with the frequencies of other axial modes. While each mode and its sidebands are originally a distinct fm oscillation, a parametric energy exchange takes place which allows the previously independent modes is coupled to other modes through the overlapping sidebands. There is then a competition between the fm oscillations for the gain of the active material, and under appropriate conditions it is possible for one fm oscillation to quench or extinguish other oscillations. The result is that the total output can be made up of only one fm carrier and its sidebands. The output signal then appears as a single frequency which is swept back and forth about line center at the modulation frequency. The fm signal E(t) can be represented as

$$E(t) = E_o \cos (\omega_c t + \Gamma \cos(\omega_m t)) \qquad (5\text{-}17)$$

where ω_c is the carrier frequency (the laser frequency) Γ is the modulation index, or ratio of the peak phase deviation to the modulation frequency, and ω_m is the modulation frequency (approx. the axial mode spacing). The relative amplitudes of the output laser modes have Bessel function relationships to each other; the central mode will have an amplitude given by $J_o(\Gamma)$, the first sidebands $J_1(\Gamma)$, and so on.

The modulation index can be written as[12]

$$\Gamma = \frac{1}{\pi} \frac{\Delta\Omega}{\Delta\nu} \delta \qquad (5\text{-}18)$$

where $\Delta\Omega$ is the frequency separation of the modes, $\Delta\nu$ is the difference in frequency between $\Delta\Omega$ and the modulation frequency ν_m, and δ is the single pass phase retardation of the phase modulator, in radians. Typical values for a He-Ne laser with $\Delta\Omega$ equal to 150 MHz are $\Delta\nu$ = 150 kHz, δ = 0.01 and Γ = 3.19.

The above equation makes it clear why fm operation is not obtained when the modulation frequency exactly equals the mode spacing. In this case, $\Delta\nu$ would be zero and Γ would be infinite. At that point the output no longer resembles an fm signal, and a new phase-locked coupled mode solution occurs.

The fm region of operation offers promise as a means to stabilize the amplitude of a cw laser. In addition, by employing the "supermode" technique or selective output coupling as discussed below, the fm laser may yield single frequency operation.

Mode Coupling in the Phase-Locked Region. When the modulation frequency is equal to the separation between axial modes ($\Delta\nu' = \Delta\Omega$), the modulation index describing fm operation becomes infinite. Under these conditions, all the modes oscillate in phase and the amplitudes of the modes have a Gaussian distribution about the center frequency. The output in this case is the same as is observed for loss modulation; that is, a train of narrow pulses with a repetition rate corresponding to the frequency separation of the modes.

Single Frequency Operation of Mode-Coupled Lasers

The "Supermode" Laser. The so called "supermode" laser[14] approach uses the controlled spectral output of the fm laser to produce a single frequency. To accomplish this, the output of the fm laser is passed through a second phase modulator located outside of the cavity which is driven at the same frequency as the internal modulator. If Γ' is the modulation index of the external modulator and ϕ is the difference in phase between the two modulators, then the output of the external modulator will be

$$E(t) = E_o \cos \left[\omega_c t + \Gamma \cos \omega_m t + \Gamma' \cos(\omega_m t + \phi) \right].$$

when Γ' is made equal to Γ and $\Phi = 180^{o}$, then $E = E_o \cos \omega_c t$. This is a monochromatic signal at a frequency near the center of the original free-running spectrum. Briefly then, the supermode laser produces a single-frequency output by first controlling the free-running modes in a specific manner through the f-m laser technique, and then converting this controlled signal to a single frequency.

The major limitation of this approach is the difficulty in building a practical external modulator for which Γ' can be equal to Γ. Typical values of Γ for an fm laser lie in the range of from 1 to 7. Because of the multiple pass characteristic of the laser cavity (as reflected in the term $\Delta\Omega/\Delta\nu'$ in Eq. (5-18), Γ may be obtained using values of δ less than 0.3. For an external modulator, $\Gamma' = \delta'/2\pi$ so that the required values of δ' are in the range of 5 to 40 radians. It is not practical at this time to construct modulators having such large phase retardations. Future development of multipass modulators using new material such as KTN or $LiNbO_3$ may make it practical to use the supermode technique to obtain a single output frequency.

Frequency Selective Output Coupling to the Mode - Coupled Laser. The important property of a mode-coupled laser utilized in this approach to obtaining single frequency operation is that all the modes are coupled together in amplitude and phase. If the gain or loss of any one of the coupled modes is changed, the relative amplitudes of the modes will still be very nearly maintained and the oscillation level will adjust such that the net average power absorbed or dissipated by all modes remains zero. The method of single frequency operation discussed here is based on the fact that there is an optimum output coupling (mirror transmission) which allows the maximum power to be taken from the mode-coupled laser- and that whether this coupling is provided as a sum of equal increments to all modes, or instead is provided entirely to one mode, is not of significance. Theory[15] has shown that if it is desired to extract the entire output as a single frequency from the 9^{th} mode from line center, the ratio of necessary

coupling to that mode, as compared to the coupling which should optimally be seen by all modes, is $1/J_q^2(\Gamma)$.

The method which has been used thus far,[15, 16, 17] to obtain the selective output coupling to a single mode of a phase modulated laser is to replace one end mirror of the laser with a Fabry-Perot etalon having a free spectral range greater than the line-width of the transition. Fine tuning of the eatlon allows the narrow transmission curvc of the etalon to be tuned so as to couple to just one mode. Ideally it should be possible to extract the total power which was originally available from all the modes to just one mode.

Single frequency operation of a multimode laser using internal modulation to achieve mode coupling and an output etalon to select a single frequency has been demonstrated for He-Ne[16] and AII lasers[17]. A He-Ne laser which as a free-running oscillator produced 68mW yielded 53mW at a single frequency using these techniques. For argon, 45mW of single frequency power has been obtained from a 100 mW free-running laser. This approach is still being refined, and higher powers can be expected, especially from argon ion lasers.

5.3.4 Laser Stabilization

Introduction. Often the laser is given all the characteristics of the more familiar high-frequency and microwave oscillators. For all but the most simple communications systems the "raw" laboratory laser does not possess the frequency stability, amplitude stability, or spectral purity equivalent to conventional sources. In the discussion that follows it is assumed that the problems of amplitude stability and spectral purity are solved; and effort will be concentrated only on some of the problems of obtaining improved frequency stability and the results obtained to date.

To first order, the oscillation frequency of a laser is determined by the mechanical properties of the optical cavity rather than the parameters of the atomic line. Thus a brief review of optical cavity properties and materials will be given. Following this is a discussion of electronic stabilization methods, and the most recent results.

System Requirements. Consider the simplest microwave heterodyne receiver with an operating frequency of 10 Gcs and an i.f. bandwidth of 1 Mcs, the required local oscillator (and transmitter oscillator) stability must be better than 10^6 in 10^{10} or 1 part in 10^4 to keep the received signal within the receiver passband. For most practical communications systems the actual stability required might be two orders of magnitude greater than this, or 1 part in 10^6. For the present, however, we will consider only the crudest system that requires only that the signal remain within the passband. The same bandwidth used in an optical heterodyne receiver, with operating frequency of 500 Tcs (Terracycle = 10^{12} cs), corresponding to a visible wavelength of 6,000 Å, requires an oscillator stability of 10^6 in 5×10^{14} or 2 parts in 10^9. By translating the same information bandwidth from the microwave to the optical region, we have changed the stability requirement from that of everyday hardware to that of a primary frequency standard. Figure 5-22 shows the fractional stability required as a function of the desired absolute stability (equal to the information bandwidth in the example considered here) for the most commonly used gas laser transitions. For a practical heterodyne system, the required stability would be an order of magnitude tighter. The curves end at the points marked "Doppler widths" since at least this stability is required to keep a given cavity mode within the gain line half-power points and to keep the laser oscillating.* Also shown for comparison with conventional sources are curves for 1-mm and 1-cm wavelengths.

The best results obtained to date[18] are shown as dots on the various curves. No stabilization has been reported for those wavelengths with no dot shown.

*Typical laboratory lasers have stability much worse than this; however, because of the multimode nature of the cavities, the dropping in and out of oscillation of a given mode is not ordinarily seen. For the optical heterodyne we must consider only single mode (i.e., single frequency) oscillators, so that the stabilization must be at least good enough to keep this mode within the gain line width.

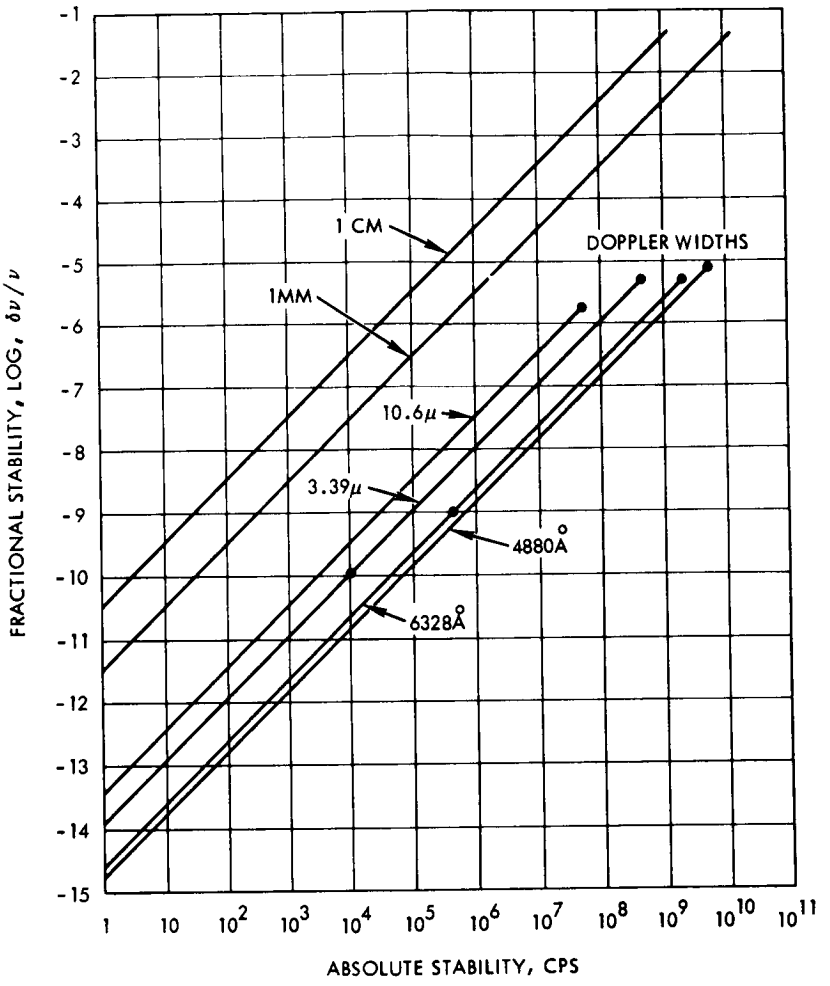


Figure 5-22. Fractional stability required as a function of desired absolute stability.

Properties of the Optical Cavity. The resonant frequency, ν, of an optical cavity formed by mirrors of radii b_1 and b_2, spaced a distance, L, apart is[19]

$$\nu = \frac{c}{2L}\left[q + \frac{1}{\pi}(1+m+n)\cos^{-1}\left\{\sqrt{(1-\frac{L}{b_1})(1-\frac{L}{b_2})}\right\}\right] \qquad (5\text{-}19)$$

where c is the velocity of light and m, n, q are three integers that describe the particular optical mode (field distribution); q is the number of half wavelengths between mirrors and m and n are number of zeros (or phase reversals) the electric field contains in the transverse direction. Typically q is of the order of 10^6, while m, n are small (0, 1, 2, . . .). Modes are usually designated TEM_{mnq}; the most useful, i.e., the ones providing the closest to a uniform illumination and thus possessing the lowest diffraction spread for a given diameter, are the TEM_{00q}. In all that follows we will assume that only these modes are present, selected by intracavity apertures, cavity dimensions, etc. In addition, we will assume that by proper choice of L compared to the gain line width, only one particular value of q will oscillate at a time.

For TEM_{00q} modes equation(5-19) reduces to

$$\nu = \frac{c}{2L}\left[q + \frac{1}{\pi}\cos^{-1}\left\{\sqrt{(1-\frac{L}{b_1})(1-\frac{L}{b_2})}\right\}\right] \qquad (5\text{-}20)$$

Since $q >> 1$, this formula, for first order effects, reduces further to

$$\nu \approx \frac{cq}{2L} \tag{5-21}$$

We see that, to first order, the cavity frequency depends only on the spacing L. The variation in frequency $\delta\nu$ caused by a change in cavity length δL is

$$\delta\nu = -\frac{cq}{2L^2}\delta L = -\frac{\nu}{L}\delta L \tag{5-22}$$

or,

$$\frac{\delta\nu}{\nu} = -\frac{\delta L}{L} \tag{5-23}$$

Second order corrections arise because of the occurance of L in the second term in equation (5-20), but equation (5-23) will be accurate enough for our purposes.

If, in addition to changes in the physical length, δL, there is a change in the effective length due to a change, δn, in the index of refraction of the intracavity medium, then

$$\frac{\delta\nu}{\nu} = -\left(\frac{\delta L}{L} + \frac{\delta n}{n}\right) \tag{5-29}$$

Equation (5-24) is most applicable to solid-state lasers where the mirrors are coated directly on a dielectric rod of index n. For gas lasers in which the active medium has index ℓ, but the atmosphere fills the space from the end windows to the mirrors, a fraction $L-\ell\ /L$, where ℓ is the length of the discharge tube, then

$$\frac{\delta\nu}{\nu} = -\left(\frac{\delta L}{L} + \frac{L-\ell}{L}\frac{\delta n}{n}\right) \tag{5-25}$$

Methods of Stabilization

Mechanical and Thermal Stabilization. The most direct method of stabilization is to control the environment of the optical cavity sufficiently well so that the resonant frequency variation falls within the prescribed limits. This requires accurate control of the cavity temperature (often in the presence of a strong heat source, for example, the laser discharge tube), control of or isolation from mechanical stresses and vibrations, and isolation from air currents (or other fluctuations in the intracavity medium.)

The change in frequency with change in cavity length caused by thermal temperature variation is given by equation (5-23) and the definition of thermal expansion coefficient, μ.

$$\frac{\delta \nu}{\nu} = -\frac{\delta L}{L} = -\mu \delta T$$

Figure 5-23 shows the fractional stability resulting from a given temperature fluctuation, δT, for several common materials suitable for cavity fabrication. This figure may be combined with Figure 5-22 to see the resulting absolute frequency change for the different laser wavelengths. To achieve a long term stability of 1 part in 10^9 using the lowest expansion coefficient material, fused quartz, a temperature control of 10^{-3} degrees Centigrade is required. By choosing a proper geometry and composite material for the package laser cavity, one may temperature-compensate to achieve perhaps an order of magnitude improvement over simple cavity expansion. No results on such experiments have been published, however. Also, there are now some experimental glasses and ceramics that are internally compensated in their composition to achieve lower expansion coefficients but, again, no results have been published.

The change in frequency produced by a change in air pressure depends on the fraction $L-\ell/L$ of the optical path that is exposed. A

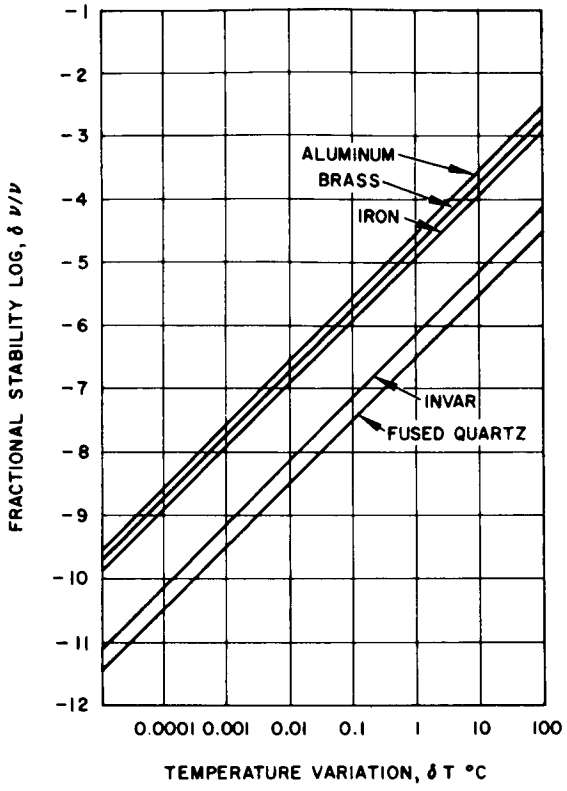


Figure 5-23. Fractional stability resulting from a given temperature fluctuation, δT, for several common materials suitable for cavity fabrication.

value of 0.1 for this fraction is typical for Brewster-angle gas lasers. The fractional frequency change produced by a change in air pressure, δp, is then

$$\frac{\delta \nu}{\nu} = - \frac{L-\ell}{L} \frac{\delta n}{n}$$

$$= - \frac{L-\ell}{L} (0.00029) \frac{\delta p}{760} \qquad (5\text{-}27)$$

$$\cong - \frac{L-\ell}{L} \times 3.8 \times 10^{-7} \delta p$$

or,

$$\cong 4 \times 10^{-8} \delta p \qquad (5\text{-}28)$$

for $L-\ell/L = 0.1$, and δp in Torr*. The long-term variation in pressure due to weather changes may be of the order of 20 to 30 Torr or more. Thus, $\delta \nu/\nu > 10^{-6}$, determined by atmospheric variations. Launching a gas laser into space ($\delta p = 760$ Torr) would produce a $\delta \nu/\nu$ of approximately 3×10^{-5}, which is greater than the maximum variation allowed to maintain oscillation within the doppler linewidth, for all the gas lasers listed in Figure 5-22. Of course, internal mirror lasers do not suffer from pressure effects, except through possible mechanical stresses set up by changes in pressure.

Perhaps the most difficult-to-describe source of laser instability is that resulting from room microphonics and mechanical vibration. Needless to say, every precaution must be taken to isolate the laser form sources of vibration. (This may be easier to do in a space vehicle than it is on the earth's surface.) Even when electronic stabilization is used with the laser, the microphonics determine the loop gain required and the trade-off between capture-range and sensitivity of the frequency discriminator to be used.

---

*1 Torr ≡

Feedback Systems Locked to a Discriminant. It is possible to generate a discriminant or discriminator characteristic as shown schematically in Figure 5-24 in several ways for use in a feedback frequency control system. First, the general characteristics of the system will be discussed and then the particular physical phenomena and arrangements to generate the discriminant will be discussed.

A discriminant must (1) measure the amount of frequency deviation from some desired center frequency (it is desirable but not necessary that the output amplitude be a linear function of the frequency deviation), and (2) indicate the sense of the deviation, usually by a positive or negative output. The discriminator output should also be independent of the signal amplitude. If it is not, the independence may be achieved as it is usually done in the radio-frequency case by limiting; unfortunately, optical frequency limiters are rare, ocurring only for those transitions which possess a high, saturable gain characteristic, such as the Xe I 3.508μ transition. Otherwise, the signal source must be stabilized to obtain amplitude independence for those amplitude-dependent discriminants.

Having obtained a suitable optical discriminator, the output is then fed to a frequency-controlling element in the laser—for example, to an electromechanical element controlling the mirror spacing, or to an electro-optic element controlling the effective index of a portion of the cavity. This servo loop may become quite sophisticated in its detail. Integrators to obtain zero-error characteristics may be included; multiple loops derived from different discriminants may be used to stabilize for different "terms" (i.e., short-term, long-term*), or different loops may be added to correct for mirror tilt in addition to mirror spacing, a second order effect. A typical system block diagram is shown in Figure 5-25.

It is useful to recall a few general facts about frequency discriminator systems. If a discriminant is derived from a resonance phenomena, then it is quite generally true that the discriminator curve

*By short term, we mean the low audio range to perhaps 1 kc; by long term, we mean hours to years.

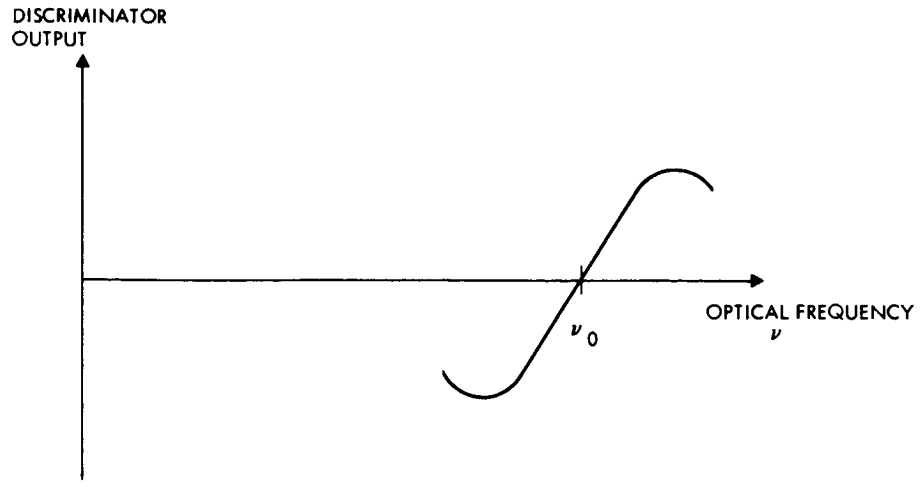


Figure 5-24. Schematic representation of an optical discriminant.

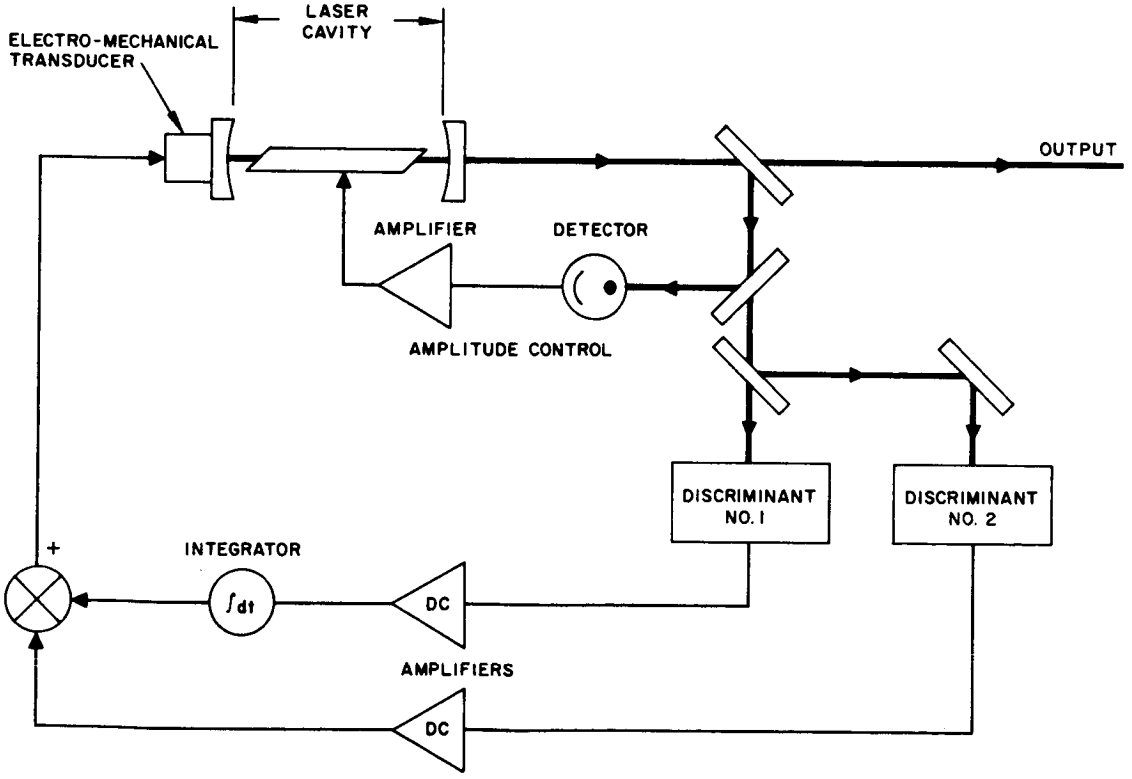


Figure 5-25. Typical feedback laser frequency control system.

will resemble the derivative of the amplitude function of the resonance. If the width of the resonance is $\Delta\nu$ then the width between the peaks of the discriminator curve will be about $\Delta\nu$; if the resonance is asymmetrical, the discriminant will be asymmetrical, etc. The finer details depend on the exact method of obtaining one from the other. It is also true that the steeper the slope of the discriminant (or the higher the gain of the feedback loop) the tighter the frequency lock will be. On the other hand, the capture range is proportional to the discriminant width.

1. Passive Cavity Discriminants

It is possible to generate an optical discriminant through interference in optical cavities. Such discriminants are probably best used for short-term stabilization, since the small cavities used can be made quite rugged and free from microphonics. However, there is no guarantee of long-term stability because of temperature drift, mechanical drift due to strains, etc.

Interference can may be classified as two-beam or multiple beam. The Michelson and Mach-Zehnder configurations are perhaps the best known examples of two-beam instruments, while the Fabry-Perot is the best known multiple beam device.

a. Fabry-Perot Interferometer Discriminator

The transmission characteristic of a typical Fabry-Perot (F-P) interferometer is shown in Figure 5-26. Typical numerical values are given in parentheses. The bandwidth $\Delta\nu$ may be only a few mc, much less than the Doppler width of a typical visible or near-IR laser transition. The simplest method of obtaining a discriminant from this resonator is shown schematically in Figure 5-27. The laser output is passed through the F-P to an amplitude detector; a bias level is subtracted from the detector output and fed in the proper phase through a dc amplifier to the mirror transducer. The resulting discriminant is shown in Figure 5-28. The output amplitude of the laser itself must be

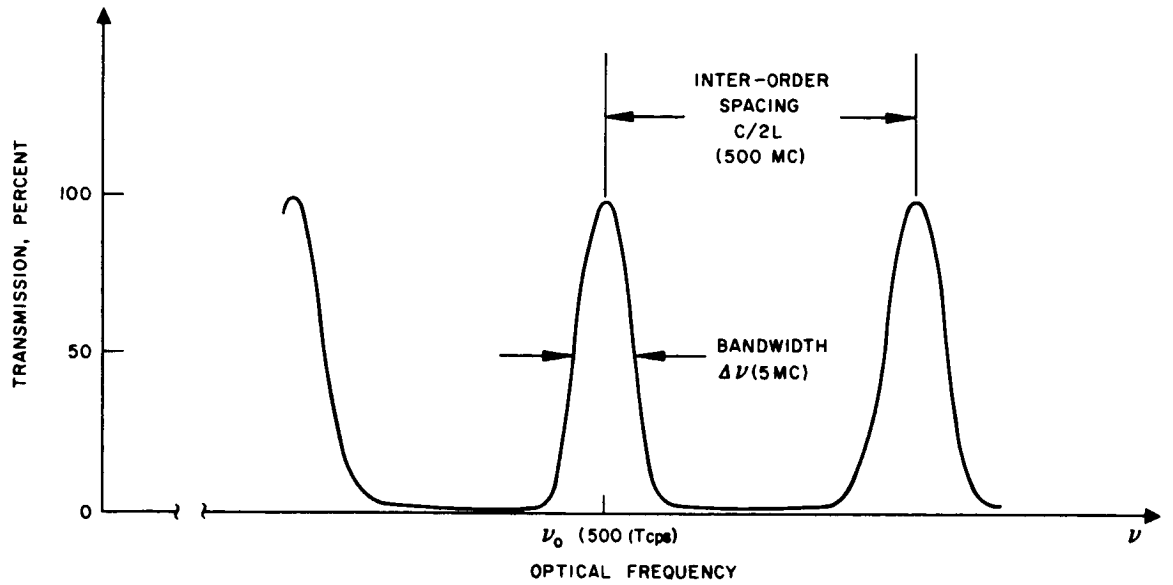


Figure 5-26. Typical Fabry-Perot transmission characteristic. Numerical values typical of gas lasers are given in parentheses.

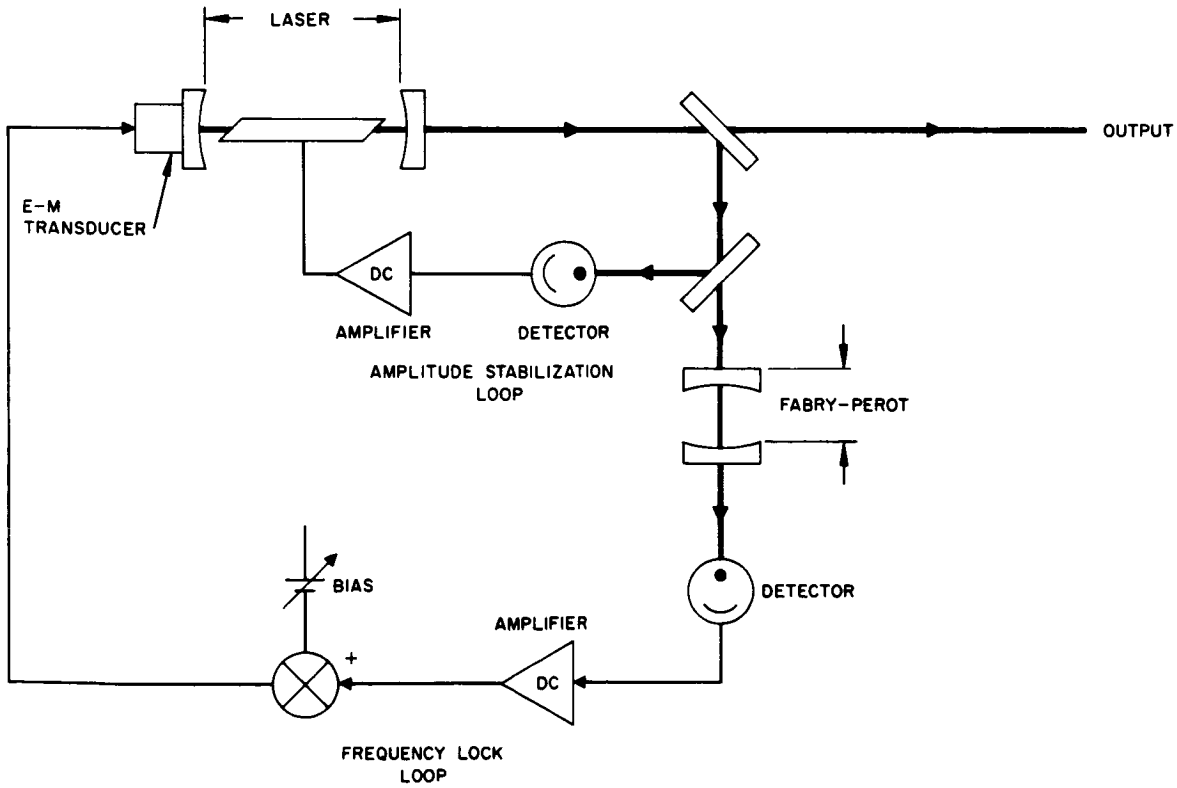


Figure 27. Simple stabilized laser using a Fabry-Perot cavity as a reference.

stabilized by an independent loop, since fluctuations in laser amplitude would be sensed the same as fluctuations in frequency, as shown in Figure 5-28.

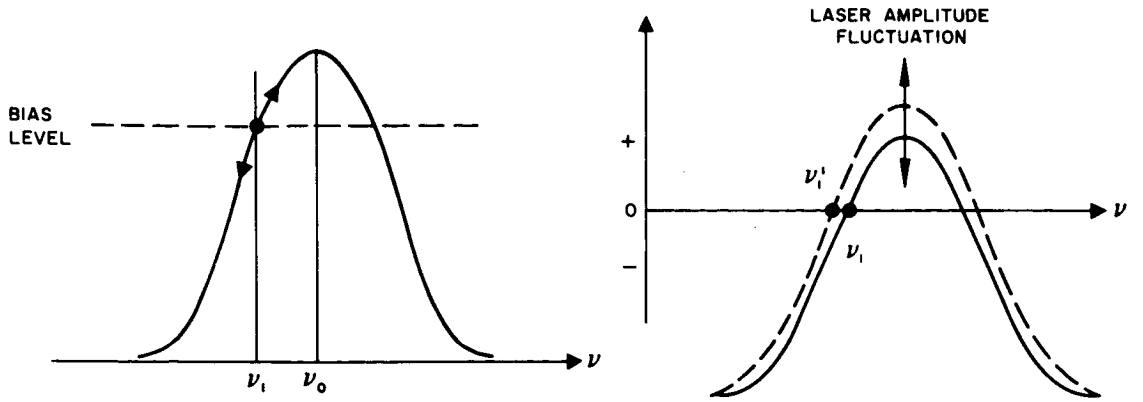
A better method is shown in Figure 5-29. Here a small modulation ("dither"), typically at an audio rate is applied to the mirror transducer. The modulated beam is passed through the F-P. A sample is envelope-detected and the modulated output of the detector is compared with the modulating source in a phase-sensitive detector. The filtered (dc) output is then amplified and fed to the mirror transducer. That the output of the phase detector produces the desired discriminant is easily seen from Figure 5-30. The amplitude of the modulation on the light increases as the slope of the transmission curve increases. The maximum modulation occurs at the inflection points ν_1 and ν_2. The modulation amplitude is zero at ν_0. (Only the second harmonic of the modulation frequency occurs at ν_0, and it is a maximum here, zero at the inflection points--it may also be used as an error signal). The phase of the modulation on the light with respect to the modulating signal also changes as the frequency passes through ν_0.

By adding a bias voltage to the dc amplifier output, the frequency may be stabilized at ν_3 rather than ν_0.

One of the disadvantages of the system shown in Figure 5-29 is that the laser output is frequency- and amplitude-modulated at the "dither" frequency. To eliminate this, the F-P mirrors may be dithered instead of the laser mirrors; the dc error signal is fed back to the laser as in the first system. The operation is otherwise the same.

b. Two-beam Interferometer Discriminator

Consider the simple arrangement shown in Figure 5-31; this system may be thought of as an unequal-arm Mach-Zehnder



(a) TRANSMISSION CURVE (EXPANDED SCALE)

(b) RESULTING DISCRIMINATOR OUTPUT. AMPLITUDE FLUCTUATIONS IN THE LASER RESULT IN FLUCTUATIONS IN FREQUENCY, ν_1 TO ν_1'.

Figure 5-28. Discriminant characteristics.

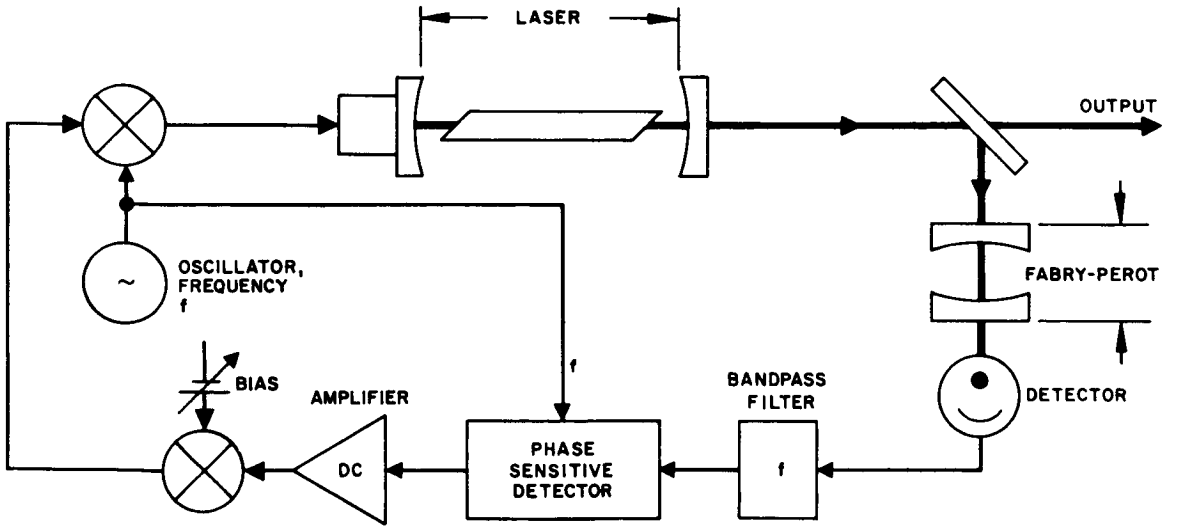
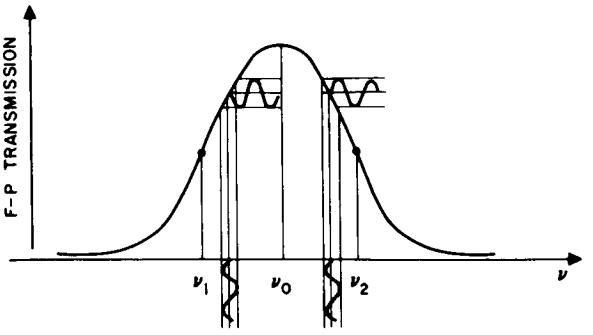
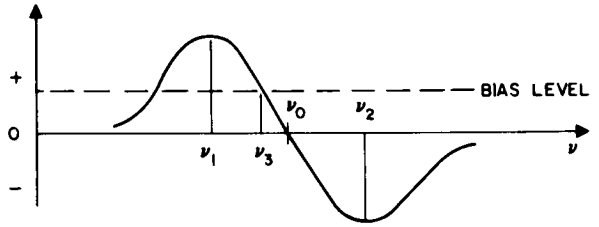


Figure 5-29. Dither method of locking a laser to a Fabry-Perot cavity resonance.



(a) TRANSMISSION OF AN F-P RESONANCE. FREQUENCY MODULATION ABOUT A POSITION OFF LINE CENTER PRODUCES AMPLITUDE MODULATION OF THE TRANSMITTED SIGNAL. RELATIVE PHASE SHIFTS 180° GOING THROUGH LINE CENTER. MAXIMUM MODULATION OCCURS AT INFLECTION POINTS ν_1, ν_2.



(b) DISCRIMINANT PRODUCED BY THIS SYSTEM. OSCILLATION IS LOCKED AT ν_0 WITH NO BIAS, AT ν_3 FOR BIAS SHOWN.

Figure 5-30. Discriminant characteristics.

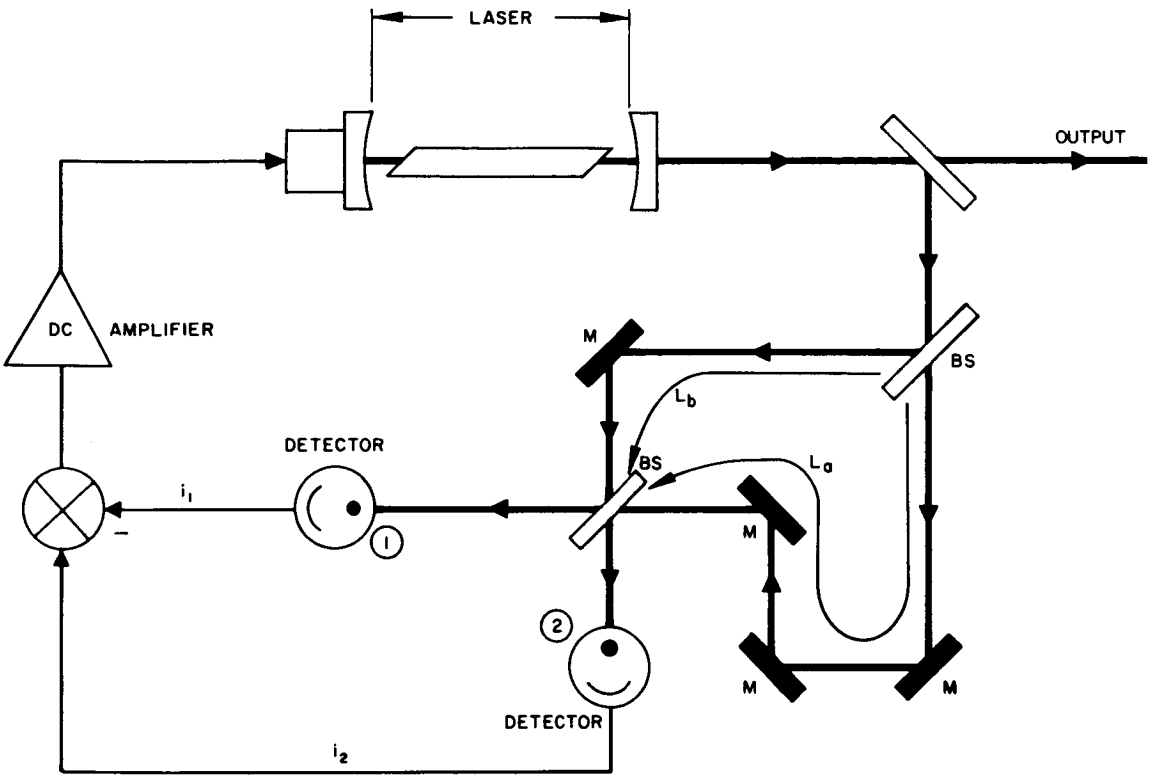


Figure 5-31. Two-beam interferometer balanced discriminator frequency stabilizing system (after Kaminow[3]).

interferometer. It is easily shown[20] that the output of photodetectors 1 and 2 is

$$\dot{Z}_1 \sim \text{SIN}^2 \left[\frac{\pi \nu}{C} (L_a - L_b) \right] \tag{5-29}$$

$$\dot{Z}_2 \sim \text{COS}^2 \left[\frac{\pi \nu}{C} (L_a - L_b) \right] \tag{5-30}$$

The photo currents may be subtracted to form the discriminant shown in Figure 5-32. The lengths L_a and L_b are adjusted so that the zero occurs at the desired optical frequency. The width of the discriminant between peaks is

$$\Delta \nu = \frac{C}{2 (L_a - L_b)}$$

and can be made smaller (more sensitive) by increasing the inequality in the arm lengths. Of course, the mechanical stability becomes worse the larger $L_a - L_b$ is made, so that there will be an optimum design for a given set of construction techniques and environment.

Instead of using two arms of physically different lengths, it is possible to use a birefringent crystal to obtain the necessary two paths of different optical length. A discriminator based on this idea has been proposed by Harris[21].

2. Atomic Line Discriminators

The laser transition itself may be used as a resonance to which the oscillation frequency may be locked. For gaseous lasers operating in the visible and near-IR, the Doppler width is much greater than the natural linewidth $\Delta\nu_N$, and much greater than widths easily obtainable with F-P cavities. Consequently, the discriminant derived from the atomic lines are broad by comparison with the cavity schemes so that the locking error is proportionally larger. However, the center frequency of the discriminant is determined by atomic parameters

rather than physical dimensions, and is thus not subject to long-term drift. These discrimants are thus better for long-term stabilization.

a. Laser Gain Curve

The Doppler-broadened gain curve of the laser medium itself may be used as the atomic resonance from which the discriminant is derived. Consider the system shown in Figure 5-27, but without the external F-P cavity. As the laser mirror is tuned across the gain curve the output will vary as shown in Figure 5-33 (This picture is highly schematic, since the output power is a monotonic function of the gain, but not necessarily a linear function.) It is assumed, of course, that the laser oscillates in only one longitudinal cavity mode across the entire line width, which requires $c/2\ell > \Delta\nu_D$. The servo loop may be closed as shown in Figure 5-27, with the laser operating on the side of the gain curve at a point determined by the bias voltage. As before, the laser amplitude must be stabilized independently.

In like fashion, the system of Figure 5-29 may be used as shown but with the external F-P removed. The curve of Figure 5-30 is now interpreted as the power output curve of the laser. The dithered signal now seeks line center (if no bias is used) or some other point on the curve determined by the bias voltage.

One new effect arises to complicate this use of the laser transition as the atomic resonance—this is the partial saturation of the medium by the oscillation, termed "hole-burning" by Bennett[22]. The small-signal gain curve for a gas laser has the simple gaussian shape shown in Figure 5-34a. The gaussian shape results from the thermal (Maxwellian) distribution of velocities of the atoms in the medium. One may think of the line width for atoms at rest being much narrower, $\Delta\nu_N \ll \Delta\nu_D$, but that groups of atoms with the same velocity are superposed with their resonances translated by the Doppler shift for that velocity. The contribution to the resulting line shape from each group is weighted by the number in that group. The presence of a strong oscillation in the medium at frequency ν_1 saturates the gain in a frequency interval $\sim\Delta\nu_N$ about it. This width, $\Delta\nu_N$, is also called the homogeneous

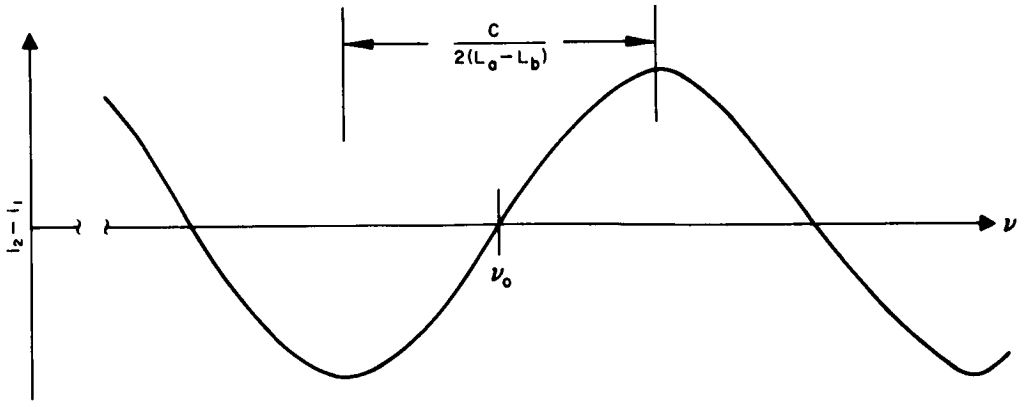


Figure 5-32. Optical discriminant obtained from a two-beam interferometer with path length difference of L_a-L_b.

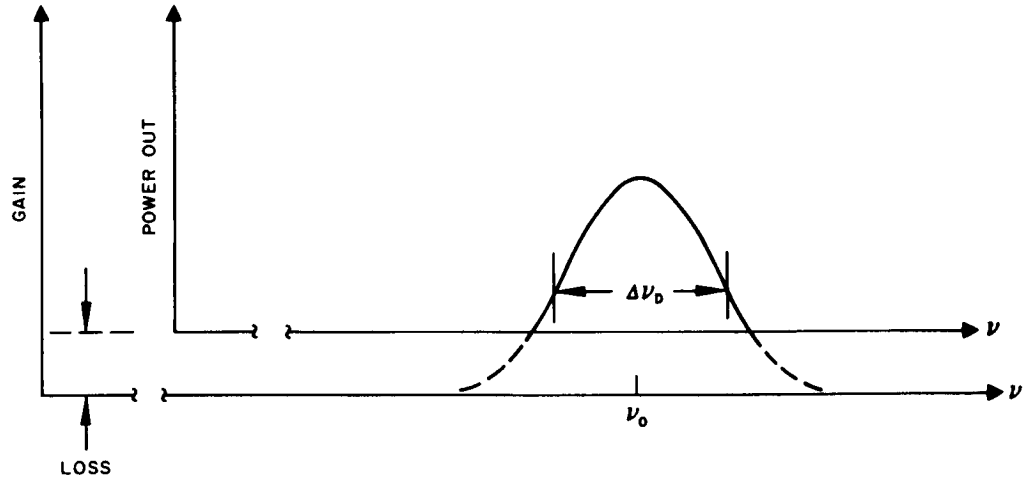
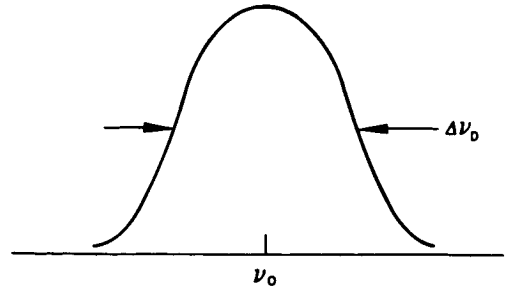
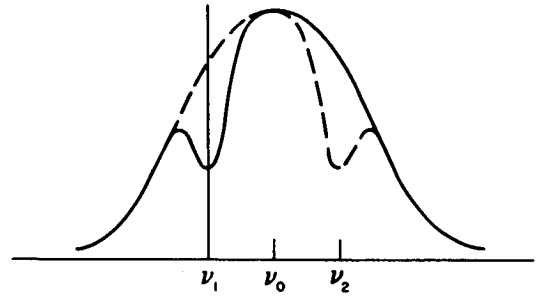


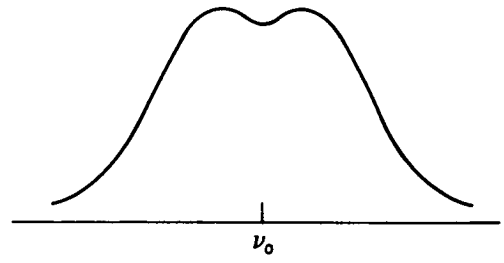
Figure 5-33. Schematic representation of the laser transition gain and power output profiles, showing line center, ν_O, doppler width, $\Delta\nu_D$, and loss line which determines threshold of oscillation.



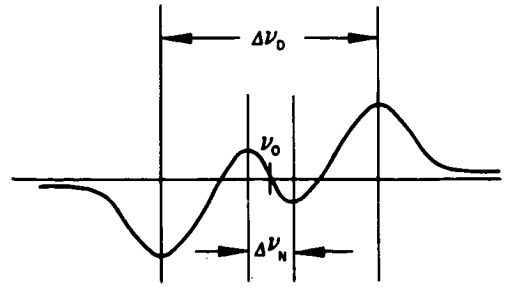
(a) SMALL-SIGNAL GAIN CURVE (GAUSSIAN)



(b) GAIN CURVE WITH HOLE CAUSED BY OSCILLATION AT ν_1 AND IMAGE HOLE AT ν_2



(c) RESULTING POWER OUTPUT CURVE, SHOWING THE LAMB DIP



(d) DISCRIMINANT DERIVED FROM (c) ABOVE

Figure 5-34. Discriminant derivation.

linewidth and represents the frequency interval over which a wave can interact with the atoms. (Clearly, a single-frequency laser does not interact with all the atoms in the laser medium.) Oscillation at ν_1 not only burns a "hole" centered at ν_1, but also a "hole" an equal distance the other side of line center, as shown in Figure 5-34b. This is easily seen if one thinks of the abscissa of Figure 5-34b as atomic velocity rather than frequency, with zero velocity occurring at ν_0. In a two-mirror laser, the oscillation at ν_1 consists of waves traveling in both directions, so that one hole is burned by one traveling wave, and the other by the same wave on the return trip. The power output thus is twice that obtained if only one traveling wave were present (e.g., in an unidirectional ring resonator). However, as ν_1 approaches ν_0, the two holes begin to overlap (i.e., the two waves are interacting with some of the same atoms, namely those nearly at rest). When $\nu_1 = \nu_0$ the two holes overlap completely, and power output is at a minimum. The resulting power output versus frequency curve is shown schematically in Figure 5-34c. The depth of the minimum depends on the strength of oscillation and gain of the medium; the width is of the order $\Delta\nu_N$. This minimum, called the Lamb Dip, is not always easy to see because of broadening by isotope mixtures. It may appear only as a flattening of the gaussian shape. The optical discriminant obtained appears as in Figure 5-34d. Locking can be maintained at three points on the curve (by appropriate phasing of the servo output.) The central slope depends on the depth of the minimum, which may be quite shallow, thereby producing a very flat and undesirable discriminant.

When the laser itself is used to generate the discriminant by dithering, it is no longer possible to keep the modulation off the output signal. However, the same transition may often be excited in an external absorption or amplifying cell. If this is possible, that cell may be modulated instead of the laser. Such a scheme is treated in the next section.

b. Zeeman Effect Methods

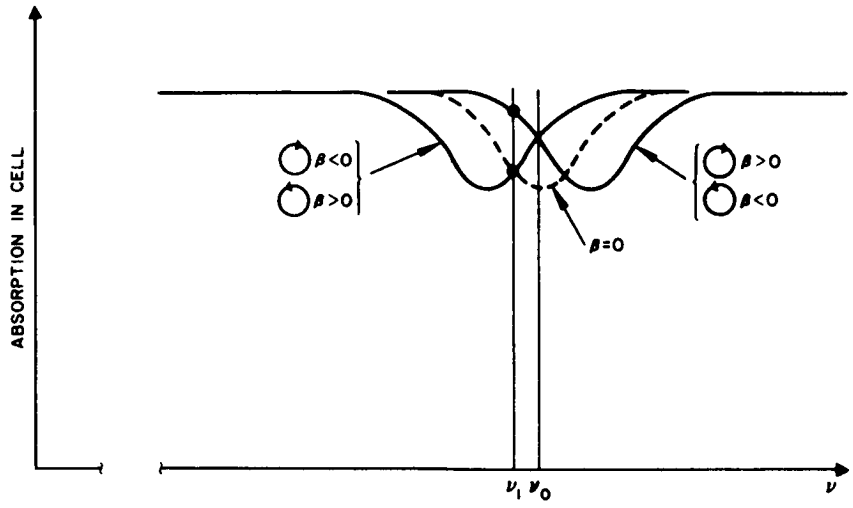
In this method[23], the discriminant is derived from the differential absorption of light passed through an external absorption cell to which is applied a square-wave modulated magnetic field. The magnetic field resolves the transition into two Zeeman componants, with the absorbtion curve for circularly polarized light of one sense shifter up for $B > 0$ and down for $B < 0$, and conversely for the other sense of circular polarization, as shown in Figure 5-35a. It is easy to see that the phase comparison of the output of the detector with the modulation will produce the desired discriminant, as shown in Figure 5-35b. The magnetic field magnitude that produces the optimum descriminant is just that which splits the line by about one Doppler width, so that the discriminant width is again about one Doppler width between peaks. The system shown in Figure 5-36 is actually only one variation. The switched magnetic field may be replaced with a dc field and an electro-optic quarter-wave plate used instead to switch from one sense of circular polarization to the other. Alternatively, the one sense of circular polarization may be passed through the cell in two oppositely-directed paths and the outputs of two detectors subtracted. This latter method requires careful balancing of the losses in the two paths, however.

It is also possible to use the circular birefringence induced by the Zeeman splitting rather than the circular dichroism, as above. A system based on this principle has been proposed by Tobias, et al[24], but not yet demonstrated.

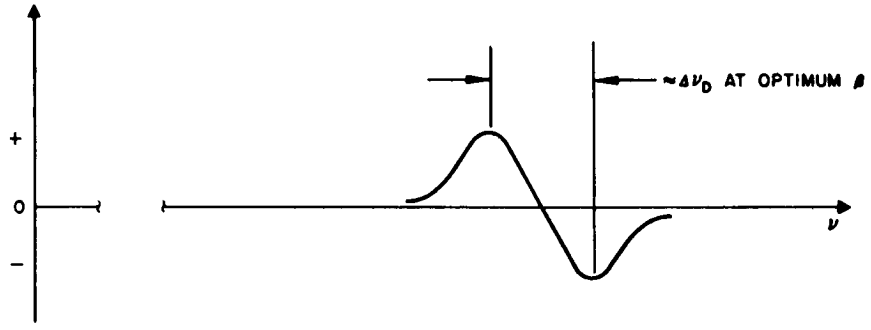
3. Other Discriminants

There are several other possible effects that may be used to derive a suitable optical discriminant. Only one, Frequency Pushing, will be considered now.

The dispersive effects of hole-burning may also be used to obtain an optical discriminant. A hole burned on one side of line center will produce a slight change in the effective index at the position of the mirror image hole and vice versa, in a direction so as to "push" the holes



(a) CIRCULAR DICHROISM IN ZEEMAN SPLIT CELL



(b) DISCRIMINANT DERIVED FROM (a)

Figure 5-35. Discriminant characteristics.

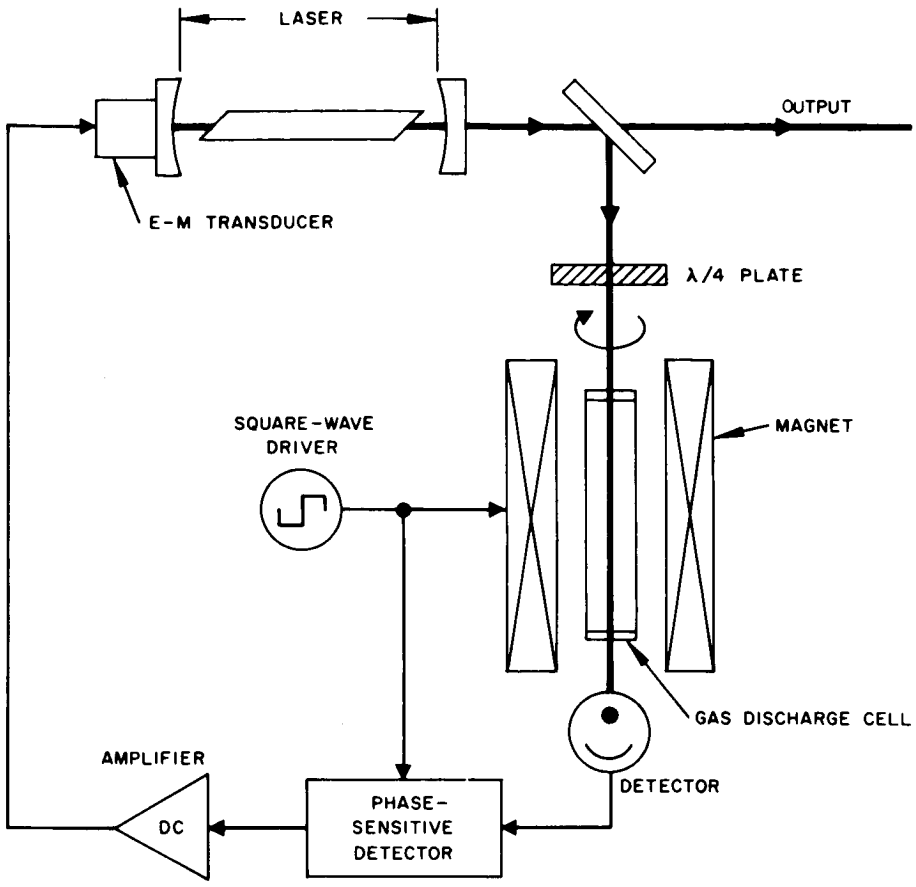


Figure 5-36. One variation of the Zeeman effect derived discriminator stabilization system.

apart. The net effect is a slight shift in the oscillation frequency that is zero at line center and of opposite sign on either side. Near line center, the strength of this effect is also approximately proportional to the gain. If one then modulates the gain (by modulating the discharge current, for example) there results a frequency modulation of the pushing effect, which changes phase by 180 degrees at line center. One advantage of this scheme is that the phase-sensitive signal to be detected is fm rather than am, and does not vanish in amplitude at line center as am dither schemes do. (Servo systems operated with error signals that vanish at the desired operating point are much more subject to noise in the physical process used to obtain the discriminant.) This fact also results in some additional complication, however, since the fm signal must be detected by heterodyning to an intermediate frequency (10.7 mc in the experimental system used by Bennett, et al) before it can be converted to an am discriminant (now in a relatively noise-free electronic discriminator). To accomplish the heterodyning, a second laser must be locked to the average frequency of the first by a slow feedback loop (afc at 10.7 mc if). The results obtained with this system, 1 part in 10^{10} for an 8-hour period seem to justify the additional complexity. For more details the reader is referred to the original paper.

State of the Art Results. For a good review and summary of results obtained to date the reader is referred to the excellent review article by White[18]. The best fractional stability for any laser system quoted is 1 part in 10^{10} for a duration of 8 hours, obtained by Bennett, et al[25], using the hole-repulsion effect with a 3.39μ He-Ne laser. For certain favorable periods of 1-minute duration or so, 1 part in 10^{12} was obtained. The best stability obtained to date for a free-running laser in a controlled mechanical and thermal environment is 2 parts in 10^{8} obtained by Collinson[26] with an rf excited 6328 Å He-Ne laser in a fused quartz cavity. Using a Brewster's angle window laser and a dither scheme, Rowley and Wilson[27] have achieved 1 part in 10^{8} at 6328 Å. Shimoda and Javan[28] have used a somewhat more elaborate

dither system with a 1.15μ internal mirror He-Ne laser to achieve 2 parts in 10^9 for many days. White[29] has used a combination of the Zeeman effect system for long-term stability and a passive cavity loop for short-term stability to achieve 1 part in 10^9 at 6328 Å.

No results of frequency stabilization experiments on the ionized argon transitions (e.g., Ar II 4880Å or Ar II 5145Å) have been reported. This is no doubt due to the difficulty in obtaining single-frequency operation with such a large Doppler linewidth (~5Gc). Intracavity methods of obtaining single-frequency operation (see discussion in Section 7.0 of this volume) show promise of self-stabilization and may well be applied to the argon ion laser, but this has not yet been accomplished.

No stabilization work has been reported yet for the 10.6μ CO_2 laser.

5.3.5 The CO_2 Laser

Theory of Operation

CO_2. Laser action in CO_2 has been observed on both the P- and R-branches of the $00^01 - 10^00$ and $00^01 - 02^00$ vibrational bands and on the P-branch of the 01^11 to 03^10 vibrational band (see references 30-48). The following discussion will be primarily concerned with the high power P-branch rotational transitions of the $00^01 - 10^00$ vibrational band which occur in a narrow wavelength interval near 10.6μ. Figure 5-37 illustrates the pertinent energy levels for CO_2 and CO_2-N_2 lasers. The latter will be discussed in the following section. For simplicity the rotational levels are not shown for each vibrational state. The vibrational levels at the left of the CO_2 energy level diagram are those of the symmetric stretching mode (ν_1), at the right are those of the asymmetric stretching mode (ν_3), and in the center are those of the doubly degenerate bending mode (ν_2).

Excitation of the upper laser level in the CO_2 laser may occur via recombination and/or cascade from higher levels, or by electron impact,

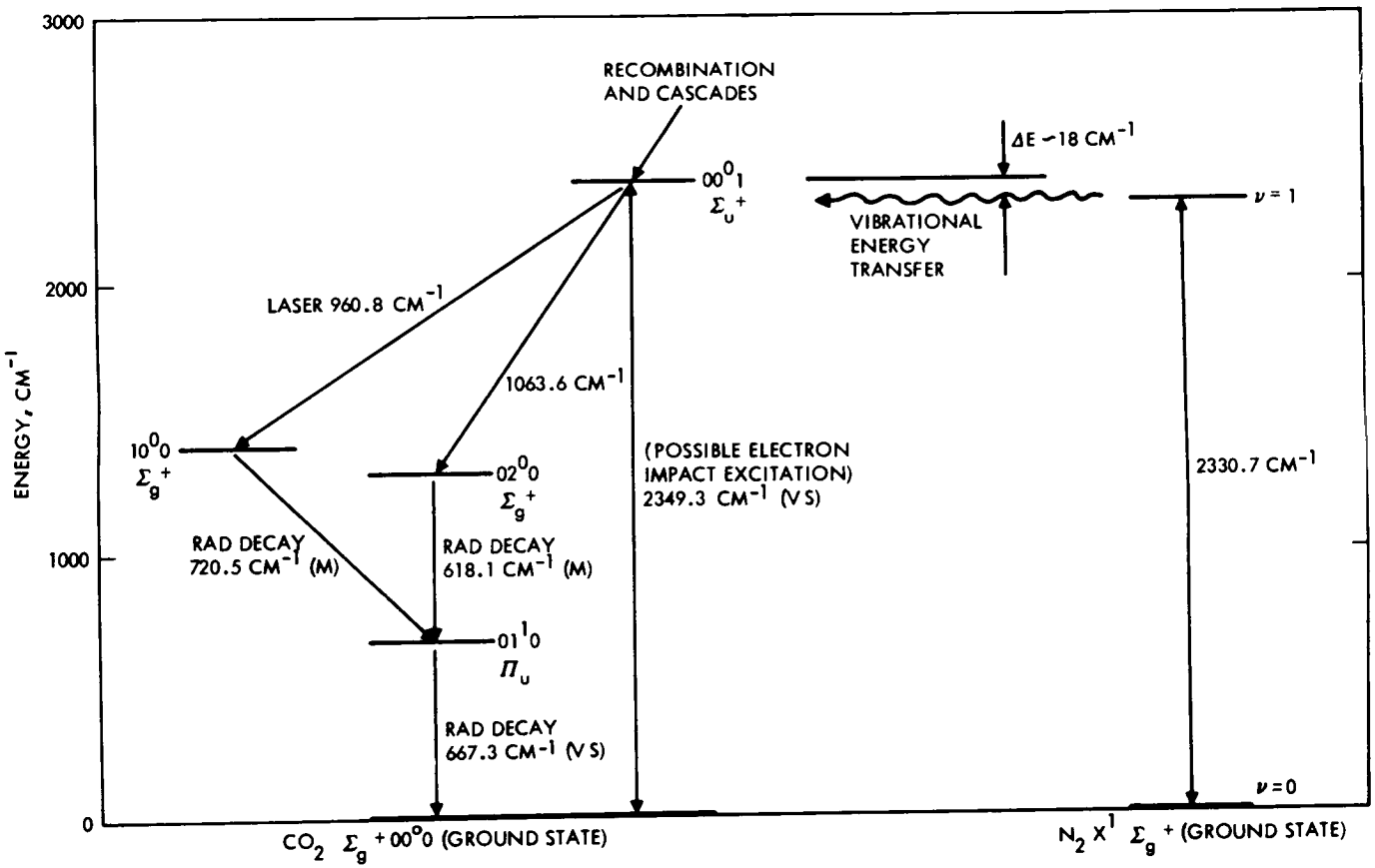


Figure 5-37. Energy level diagram showing pertinent levels in CO_2 and N_2.

either directly or via a compound. Experiments indicate that the dominant mechanism may vary as conditions of the discharge are changed.

Decay of vibrationally excited CO_2 molecules is by spontaneous emission and/or collisional de-excitation. The calculated spontaneous radiative transition probabilities between the lower vibrational levels in CO_2 are listed in Table 5-6[49].

It is evident that for transitions other than $00^0n_3 - 00^0(n_3-1)$, spontaneous radiation is negligible. The lifetime of the 00^01 level, in the presence of radiation trapping, is $\sim 2 \times 10^{-2}$ seconds.

The amount of power produced by CO_2 lasers requires that the lifetime of the lower laser levels be 10^{-3} seconds. Consequently, de-excitation by CO_2-CO_2 collisions involving the conversion of vibrational energy to translational energy is the dominant relaxation mechanism. The levels of each vibrational mode reach internal thermal equilibrium via vibration-vibration exchange then cross-relax to the fastest relaxing mode. In CO_2 the most probable vibration-translation relaxation mechanism involves vibrational quanta of the lowest frequency mode ν_2,[50-52]. i.e.,

$$CO_2(01^10) + M \rightarrow CO_2(00^00) + M + KE$$

Transition	Dominant Branch	Spont. Trans. Prob.
$00^01 - 10^00$	P	$0.34\ sec^{-1}$
$00^01 - 02^00$	P	$0.20\ sec^{-1}$
$10^00 - 01^10$	Q	$0.53\ sec^{-1}$
$02^00 - 01^10$	Q	$0.48\ sec^{-1}$
$01^10 - 00^00$	Q	$1.07\ sec^{-1}$
$00^01 - 00^00$	P	$2 \times 10^2\ sec^{-1}$

Table 5-6. Spontaneous radiative transition probabilities in CO_2.

where M is CO_2 or another species. The corresponding relaxation rates to the ground rate for molecules in the lowest vibrational levels of the ν_1 and ν_3 modes, 10^00 the lower laser level, and 00^01 the upper laser level, respectively, are several orders of magnitude slower. The lower laser level (10^00) relaxes to 01^10 by

$$CO_2\ (10^00) + M \rightarrow CO_2\ (01^10) + M + KE \qquad (5\text{-}33)$$

or, since the 10^00 and 02^00 levels are in Fermi resonance, by

$$CO_2\ (10^00) + CO_2\ (00^00) \rightarrow CO_2\ (02^00) + CO_2\ (00^00) \qquad (5\text{-}34)$$

or

$$CO_2\ (10^00) + CO_2\ (00^00) \rightarrow 2\ CO_2\ (01^10) \qquad (5\text{-}35)$$

In each case, the final relaxation process is governed by (5-32). Calculations indicate that the rates for processes (5-32) and (5-33) are of the same order of magnitude, although the latter is somewhat slower, while the rate for 02^00 - 01^10 vibration-translation relaxation is faster[51]. Thus, the de-excitation of the lower laser level 10^00 is by vibration-vibration exchange with the bending mode which then relaxes via vibration-translation exchange to the ground state. The rate of this latter process controls the relaxation of the lower laser level.

$\underline{CO_2 - N_2}$. The increase in output power from the CO_2-N_2 laser over that from the pure CO_2 discharge may be attributed to population of the upper laser level in CO_2 by transfer of vibrational energy from N_2 molecules in the $\nu = 1$ vibrational level of their electronic ground state. The selective excitation of the CO_2 molecule from its ground state to the 00^01 state takes place during a two-body collision involving a CO_2 ground-state molecule and a vibrationally excited N_2 molecule.

From Figure 5-37 it is evident that the $\nu = 1$ vibrational level of N_2 at 2330.7 cm^{-1} is in very close coincidence with the $00^0 1$ vibrational level in CO_2 at 2349.16 cm^{-1} ($\Delta E = 18.46\ cm^{-1} < kT_{rotational} \sim kT_{translational} \sim 210\ cm^{-1}$.) In addition, since N_2 has a zero permanent dipole moment, vibrational relaxation times for excited N_2 molecules are long and are determined by collisions with other molecules and walls. The addition of CO_2 allows excited N_2 molecules to relax via vibration-vibration exchange with $00^0 n_3$ levels of CO_2. The process with the largest cross section is

(5-36)

$$N_2 (\nu = 1) + CO_2 (00^0 0) \rightarrow N_2 (\nu = 0) + CO_2 (00^0 1) - 18\ cm^{-1}$$

For N_2 molecules in higher energy vibrational states

(5-37)

$$N_2 (\nu = \nu') + CO_2 (00^0 0) \rightarrow N_2 (\nu = \nu' - 1) + CO_2 (00^0 1) + \Delta E$$

Because of the large energy difference (~950 cm^{-1}) between the lower laser level $10^0 0$ in CO_2 and the $\nu = 1$ vibrational level in N_2, the cross-section for excitation of ground state CO_2 molecules to the lower laser level through collisions with N_2 ($\nu = 1$) is much smaller than that for the reaction described in equation 5-36. In addition, the excitation of CO_2 $(00^0 0)$ molecules to the $10^0 0$ energy level involves a reaction in which both transitions; i.e., $N_2(\nu=1) \rightarrow N_2(\nu=0)$ and CO_2 $(00^0 0) \rightarrow$ CO_2 $(10^0 0)$, are optically forbidden. It has been shown[53] that, for reactions involving collisions of the second kind, the cross-section for a reaction involving two optically forbidden transitions is smaller than that involving one forbidden transition for the same energy level discrepancy.

CO_2-N_2-He. A significant increase in power over that produced by CO_2 and CO_2-N_2 lasers is achieved by the addition of relatively large amounts of He to either discharge. At this time the role of the He has not been clearly defined although it has been determined

experimentally that the He affects the excitation of the upper laser level as well as the relaxation of the lower laser level. The former effect may result from the creation of plasma conditions (e.g., electron temperature) which favor excitation of the $00^0 1$ level in CO_2. In addition there is the possibility of a three-body collision involving the transfer of energy from metastable $He(2^3S_1)$ atoms to N_2 molecules[40], viz

$$He\ (2^3S_1) + 2\ N_2 \rightarrow He + 4N + \Delta E$$

where $\Delta E < 0.1$ eV. The atomic nitrogen, thus produced, recombines to form vibrationally excited nitrogen in high vibrational levels which excite CO_2 molecules to the $00^0 1$ vibrational level by the processes described in Equations 5-36 and 5-37.

The addition of He has little effect on the decay of the upper laser level $00^0 1$. However, the rate for vibration-translation relaxation of CO_2 via the bending mode is dependent on the reduced mass of the colliding particles and is thus greater for CO_2-He collisions than for those involving two CO_2 molecules. The former collision process is one to two orders of magnitude more efficient for relaxation[54]. Thus it appears that improved laser performance due to the addition of He may be due in part to an increased relaxation rate of the terminal laser level.

An entirely different role may be played by the He in that it is able to reduce the gas temperature because of its high thermal conductivity. A low gas temperature leads to a reduction in thermal excitation of the lower laser level, which may be appreciable since the energy is only 0.17 eV, as well as an increase in the gain coefficient for the laser transitions. (These effects are described more fully in Section 1) - e).

Other Gas Additives. The addition of other gases (air, CO, H_2, water vapor, etc.) influences the output from the CO_2 laser in varying degrees; however, none is as effective as He. In the case of H_2 and

water vapor it is felt that the primary effect is to enhance the relaxation rate of the terminal laser level by collisions. Air is thought to combine the roles of N_2 and water vapor. CO, which has a long relaxation time, may selectively excite CO_2 by vibration-vibration exchange with asymmetric stretching mode ν_3.

Effect of Gas Temperature. The lower level 10^00 of the laser transition lies sufficiently close to the ground state (~0.17eV) that a significant fraction of the CO_2 molecules may be thermally excited to that state if the gas temperature becomes too high. This effect may be quite appreciable in large diameter tubes. Under the assumption that laser action will cease when 10 percent of the ground state molecules are thermally excited to the lower laser level, the cut-off temperature is 662°C. With a power input of 0.1 W/cm^3 and in the absence of any cooling, this temperature can be reached and laser action cease in times of the order of 0.1 second. Under steady-state conditions in which cooling is provided at the discharge tube walls, thermal excitation of the lower laser level will prevent oscillation along the laser axis for diameters greater than approximately 5 inches.

With the assumptions that the laser transition is predominantly Doppler-broadened and that the rotational level populations are described by a Boltzmann distribution at a temperature $T_{rotational} \simeq T_{translational}$, it can be shown [31, 41] that the gain coefficient of the laser transitions is inversely proportional to the molecular temperature, i.e.,

$$g_o \propto (T_{translational})^{-3/2} \tag{5-39}$$

The validity of the latter assumption is supported by the fact that rotational thermalization times are of the order of microseconds while in CO_2 the vibrational and radiative lifetimes are of the order of milliseconds and seconds, respectively. In addition, increased laser gain and output power have been observed experimentally by cooling the walls of the discharge tube.

Frequency Spectrum

Vibrational-Rotational Transitions. The transitions of the P and R branches of the $00^0 1 - 10^0 0$ vibrational band of CO_2 on which cw laser oscillation has been observed are listed in Table 5-7. All the lines do not oscillate simultaneously but are obtained by placing a dispersive element, such as a prism[47] or grating[45], in the optical cavity. In the absence of such a wavelength selective resonator, several of the strongest P-branch transitions will oscillate at the same time.

Single Wavelength Operation. As indicated above, the gain of the laser is sufficiently high that several rotational transitions may oscillate simultaneously. For single wave-length operation, wavelength selective cavities utilizing prisms or gratings will be required. Because of the rapid rotational thermalization ($\tau \sim 10^{-6}$ sec), molecules in the other rotational levels of the $00^0 1$ level may cross-relax into the upper laser level of the rotational vibrational transition which is selected by the resonator configuration. Similarly, molecules in the lower laser level may cross-relax into other rotational levels of the $10^0 0$ level. Consequently, the output power obtained under single wavelength operation may be close to the total power produced when several transitions oscillate simultaneously.

Single Mode Operation. The narrow Doppler width of the laser transition (50-75mc/s) allows operation in a single axial mode with resonators 2 to 3 meters long. Single transverse mode operation in low power lasers ($\leq$10W) may be achieved by proper resonator design. However in very high power CO_2 lasers, transverse mode control may be complicated by the presence of self focusing and the production of a filamentary structure in the laser output[55].

Table 5-7. Continuous wave laser oscillation wavelengths in the 00^01 to 10^00 band of CO_2

Transition	Measured Frequency (cm^{-1})	Transition	Measured Frequency (cm^{-1})
P_2	959.43	R_2	963.33
P_4	957.76	R_4	964.74
P_6	956.16	R_6	966.18
P_8	954.52	R_8	967.73
P_{10}	952.88	R_{10}	969.09
P_{12}	951.16	R_{12}	970.50
P_{14}	949.44	R_{14}	971.91
P_{16}	947.73	R_{16}	973.24
P_{18}^*	945.94	R_{18}^*	974.61
P_{20}^*	944.15	R_{20}^*	975.90
P_{22}^*	942.37	R_{22}^*	977.18
P_{24}^*	940.51	R_{24}^*	978.47
P_{26}	938.66	R_{26}	979.67
P_{28}	936.77	R_{28}	980.87
P_{30}	934.88	R_{30}	982.08
P_{32}	932.92	R_{32}	983.19
P_{34}	930.97	R_{34}	984.35
P_{36}	928.94	R_{36}	985.42
P_{38}	926.96	R_{38}	986.49
P_{40}	924.90	R_{40}	987.56
P_{42}	922.85	R_{42}	988.63
P_{44}	920.77	R_{44}	989.61
P_{46}	918.65	R_{46}	990.54
P_{48}	916.51	R_{48}	991.47
P_{50}	914.41	R_{50}	992.46
P_{52}	912.16	R_{52}	993.34
P_{54}	909.92	R_{54}	994.18
P_{56}	907.73		

*Strongest transitions in the group

Scaling Laws

Variation of Output Power. The variation of output power per unit volume and per unit length as a function of discharge tube diameter is shown in Figure 5-38. The data points which are shown represent actual reported performance of laser oscillators as achieved by Hughes, Raytheon, Perkin-Elmer, Bell Telephone Laboratories, and C.G.E. in France. The shaded areas represent an estimate of the uncertainty implied with the dark curve through the center representing a best estimate for use in establishing design criteria. Because of the scatter of data points in the 6 to 9 cm range, extrapolation of the power per unit length parameter is difficult. The power per unit volume estimates are more consistent and may indicate the existence of some asymptotic value.

Variation of Input Power. In Figure 5-39, the axial electric field along the discharge tube for optimum laser performance is plotted as a function of tube diameter. As expected, it decreases rapidly with diameter until the loss of ions to the walls becomes negligible. It is apparent that the field is approaching an asymptote of approximately 30 volts/cm. In Figure 5-40, the optimum discharge current is plotted as a function of tube diameter. Although difficult to prove conclusively from the data it is expected that this current is increasing with a dependence very close to $(\text{diameter})^2$. These conditions lead to an asymptotic value for input power of approximately 0.1 watt per cm^3 as the diameter is increased beyond 10 cm. Since high efficiency operation (i.e., approximately 18 percent) has been claimed for diameter of 6 centimeters it is reasonable to assume that an output power of 10 to 20 milliwatts per cm^3 is a possibility. This is consistent with the graphic extrapolation shown in Figure 5-38.

Modes of Operation

CW. As illustrated in Figures 5-39 and 5-40, the CO_2-N_2-He laser is a relatively low current, high voltage discharge. These

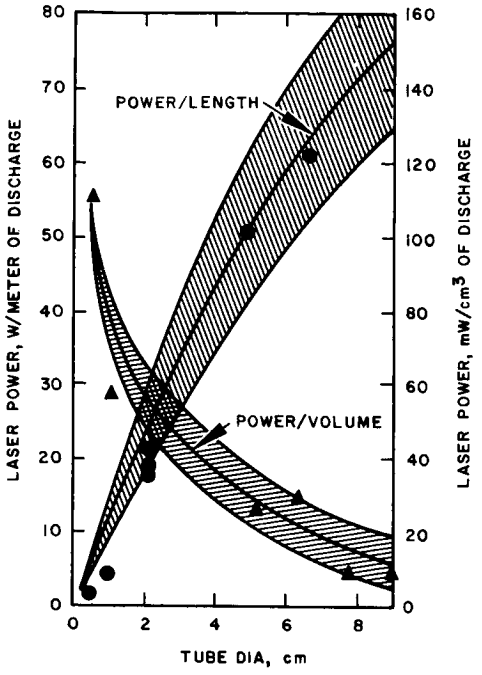


Figure 5-38. Variation of laser power per unit volume of discharge and laser power per unit length of discharge with discharge diameter.

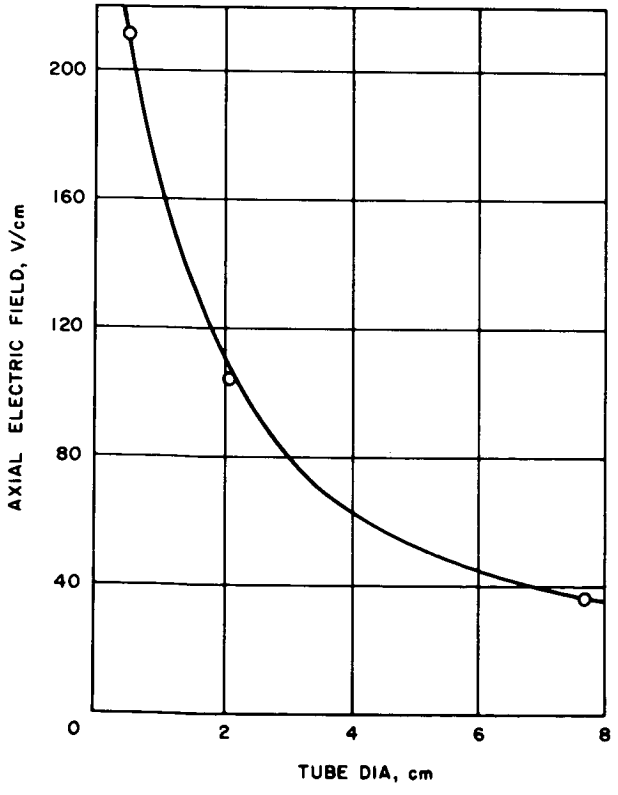


Figure 5-39. Variation of axial electric field with discharge tube diameter.

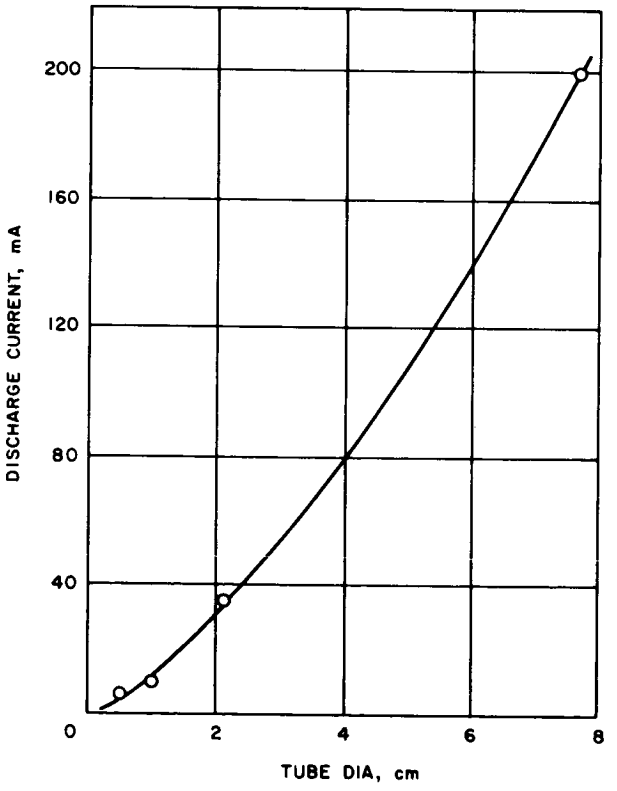


Figure 5-40. Variation of optimum discharge current with discharge tube diameter.

features, plus the fact that heated oxide coated cathodes are extremely susceptible to poisoning by this type of discharge, suggest the use of cold cathode discharges. Both dc and ac excitation can be used. Although the latter method produces slightly higher power[47], it has several disadvantages. The light output is amplitude modulated at twice the ac excitation frequency, and, because the discharge is self-striking, the current, and hence the output, exhibits random fluctuations.

By using rf excitation, the electrodes can be removed from the discharge altogether. With small diameter discharge tubes (< 2.5 cm), rf excited lasers have exhibited essentially the same output power, efficiency and pressure dependence as comparable dc excited tubes. However, the latter have two distinct advantages: (1) no rf interference in auxiliary equipment, and (2) higher power dc supplies are easier to build and operate. No data are presently available regarding rf excitation of large diameter (≥5 cm) lasers.

Pulsed. High peak power can be obtained by exciting the CO_2-N_2-He laser with dc pulses (e.g., peak power of 825 w from a laser producing 17 w cw, current pulse 100 μsec, light pulse 150 μsec[47]). No delay is observed between the current pulse and the light output.

Q-Switched. Because of the long lifetime ($\sim 2 \times 10^{-2}$ sec) and small spontaneous emission transition probability (~ 0.34 sec^{-1}) of the upper laser level 00^01, the CO_2-N_2-He laser may be Q-switched[56]. Lasers whose cw output is a few watts will produce giant pulses of the order of 10 kw with a pulse length of approximately 100 nsec when operated in the Q-switched mode. The maximum switching speed is determined by the rate at which the upper laser level is excited and is in the neighborhood of several kc/s.

State-of-the-Art in CO_2 Lasers. Since the first report of laser action in CO_2 in April 1964, CO_2 laser technology has developed at a much faster rate than that of any other laser device. In approximately two

	P_{out} (w)	Length (m)	Diam (cm)	Eff. (%)	Fig. of Merit		Comments	Organization
					P/λ^2 (w/m^2)	$\eta^2 P/L\lambda^2$ (w/m^3)		
1)	500	10	5	15.5	4.5×10^{12}	1.1×10^{10}	Multimode output Flowing gas system	Raytheon
2)	10	0.5	1	10	8.9×10^{10}	1.8×10^{9}	Single TEM_{oo} mode output Sealed-off tube	HRL
3)	150	2.4	6	18	1.4×10^{12}	1.8×10^{10}	Multimode output Flowing gas system	CGE (France)
4)	130	4	7.5	13	1.2×10^{12}	5×10^{9}	Multimode output Flowing gas system	BTL

Table 5-8. State-of-the-art CO_2 lasers (May 1966).

Active Material	Wavelength (μ)	Output Power	Dimensions of Active Material	Comments	References
1. He-Ne	0.6118 0.6328 1.084 1.152	5 mW 50 mW 5 mW 20 mW	6 mm x 1.8 m	Single mode, commercially available	57
2. He-Ne	0.6328	900 mW	10 mm x 5.5 m	research devices	58
3. He-Ne	0.6328	100 mW	5 mm x 1.2 m		
4. Xe	3.5 9.0	0.1 mW 0.5 mW	2.6 mm x 50 cm	research device	59
5. Ar^+	0.4579(0.05) 0.4765(0.1) 0.4880(0.25) 0.4965 (0.1) 0.5107 (0.1) 0.5145 (0.4)	10 W	6 mm x 60 cm	research devices, 0.1 - 0.2% efficiency	60
6. Ar^+	(as in 5)	16 W	4 mm x 2.6 m		61
7. Ar^+	0.4880	1 W	3 mm x 45 cm	airborne development device	62
8. CO_2	10.57 (0.75) 10.59 (0.25)	16 W	25 mm x 2.0 m	4.0% efficiency, single mode for each line	63
	10.59	135 W		15% efficiency	64
9. Cr^{+3}	0.6943	70 mW	2 mm x 2.54 cm	water cooled	65
10. $Nd^{+3}(CaWO_4)$	1.06	1 W	3 mm x 3.5 cm	methyl alcohol cooling (approximately 300°K)	66
11. Nd^{+++}(YAG)	1.06	1.5 W	2.5 mm x 3.0 cm		67
12. Nd^{+++}(YAG)	1.06	0.5 W	---	water cooled, commercially available, portable	68
13. $Dy^{+2}(CaF_2)$	2.36	0.75 W	4.8 mm x 2.54 cm	liquid Neon (27°K) bath	69
14. GaAs	0.84	12 W	0.5 mm x 0.4 cm (diode dimensions)	liquid He (4°K) bath, 23% efficiency	68

Table 5-9. CW laser oscillators.

years the output power has been increased from milliwatts to hundreds of watts, the latter obtained at efficiencies of 10 to 20 percent. Table 5-8 lists the characteristics of the more recently developed CO_2-N_2-He lasers. The figures of merit used are those defined in section 5.3.6.

5.3.6 Available Optical Source Performance

This section provides an initial survey of lasers for space communications and tracking. The list presented here is far from complete but does exhibit those lasers which have outstanding performance with respect to those parameters of systems relevance.

Laser oscillators have been classified according to their mode of operation as cw, pulsed - high energy, pulsed - high power, pulsed - high rep rate. Correspondingly, the following operating characteristics will be of interest: average power, pulse energy, peak pulse power, pulse repetition frequency. Information on other aspects of laser oscillators will also be required: namely, wavelength, frequency stability, beam divergence/lateral coherence, noise, efficiency, temperature, size and weight.

Cw Laser Oscillators. The first type of laser oscillator to be discussed is the cw laser oscillator. The output signal is a highly monochromatic beam of light with nearly constant output power.

The active materials which are used in cw laser oscillators include gases, solids and semiconductors. While laser action has been demonstrated in materials other than those listed in Table 5-9, only the most commonly used or most promising materials are included here.

The wavelengths available for operation comprise a set of discrete spectral lines associated with the various active materials. Generally more than one line is excited in a given laser; however, techniques exist for suppressing oscillation on all but the desired lines. Laser frequency tuning may also be accomplished, but only over a very limited range (usually less than one part in 10^4).

The output powers of existing cw laser oscillators vary from a few milliwatts for single mode operation to hundreds of watts for multifrequency operation in gas lasers. Very recently 40 watts cw has been realized in a Nd:YAG laser pumped by an experimental 20 kw argon lamp. The most important factor limiting the output power is the very low energy conversion efficiency of most cw lasers. It usually ranges from 0.01 percent to 1.0 percent; however, a few exceptions are the CO_2 molecular laser (15 percent) and the GaAs injection laser (approximately 25 percent at cryogenic temperatures). Scattering losses in the laser material and in the cavity mirrors must be overcome in order to obtain high power output in practical devices. In addition, cryogenic temperatures are needed in present injection lasers for efficient operation.

The spectral bandwidth of the laser output, which is an indication of the frequency stability, may range from a few cycles per second to several tens of gigacycles per second. Thermal fluctuations and mechanical vibrations are chiefly responsible for line broadening in single mode operation. For multimode operation, the spectral width of the output is closely related to the fluorescence linewidth of the atomic transition. Gas lasers exhibit the most monochromatic output since their fluorescence linewidths are typically one to two orders of magnitude less than for solids or semiconductors.

The spectral characteristics of laser oscillators depend primarily upon the material used. As mentioned previously, gas lasers display a more narrow fluorescence linewidth than solid or semiconductor lasers. The linewidths of these latter oscillators vary from between 0.001Å for ruby at cryogenic temperatures to several hundred Å for Nd:glass. It should be pointed out that the linewidth of ruby oscillators depends heavily on the operating temperature.

The beam divergence for axial mode operation is given in the diffraction limit by λ/D, where λ is the operating wavelength and D is the diameter of the beam at the output aperture. For most laser oscillators, the output beam diameter is of the order of a few millimeters, so beam

divergences of the order of 0.1 milliradians can be obtained for visible light. (The semiconductor laser oscillator is somewhat unique since the active region has dimensions of the order of a few microns. The resulting beam divergence, even in the diffraction limit, is of the order of a few degrees.) For maximum output power, however, non-axial modes are excited in the oscillator and the beam divergence increases to as much as 10 milliradians.

The beam divergence of the output from a laser oscillator can be made smaller by passing the beam through an optical system. The price one must pay for smaller beam divergence is a larger beam diameter and a small loss in signal strength due to reflection and absorption in the lens system.

The noise in laser oscillators can be of three types: 1) spontaneous emission noise, 2) gain fluctuations, and 3) mode-interference noise. Except for operation near threshold, spontaneous emission noise can be neglected. Gain fluctuations due to pump power modulation can generally be reduced to an insignificant level by careful design of the pump source and associated power supplies. Mode-interference noise is not so easily taken care of. It does not occur in lasers operating in a single mode; however, if two or more modes are excited, then beat frequencies between the various modes will be produced. These may range from several kilocycles per second to hundreds of megacycles per second. Schemes for phase locking the various modes by means of an intracavity modulator are currently under development and promise to make it possible to obtain high power output with little or no mode-interference noise.

Laser operation is frequently compromised by thermal problems. One of these is associated with the low energy conversion efficiency which leads to excessive heating of the laser components. Another problem is the reduction of optical quality of the laser material due to thermally-induced distortion, or stress birefringence. The effect of a non-linear temperature distribution across the laser rod, such as that which arises during the pump cycle, is to cause depolarization of the

laser output. This depolarization is not constant but displays a radial dependence. As a result, the output from the laser has a seemingly random polarization and, if one requires a polarized signal, the beam must be repolarized, introducing a loss in signal energy.

Pulsed Laser Oscillators

Pulsed Laser Oscillators - High Energy. High energy pulsed laser oscillators produce a short burst of intense monochromatic radiation in a narrow output beam. These lasers operate in a quasi-cw conversion mode - converting incoherent flashtube illumination into laser radiation. Pulse length is limited only by thermal considerations. The pulse may be smooth, as in the case of the super-fluorescent radiator, or it may consist of a series of closely spaced, random or regular spikes. Pulse repetition frequencies are very low (about one per minute for energies measured in tens of joules) due to the time required for cooling of the laser components.

Data for some high energy laser oscillators are given in Table 5-10. Indicative of the energy available, an off the shelf pulsed ruby oscillator with a guaranteed energy of 140 joules is available. These materials are used because of their relatively high ruggedness and conversion efficiency when used to convert flashtube pumping radiation to laser emission. Large size rods of good optical quality are available to match large pumping flashtubes.

Material	Wavelength (μ)	Pulse Energy (J)	Pulse Duration	Rod Size	Comments	References
1. Ruby	0.6943	50	0.5 msec	12 mm x 15 cm	commercially Available; 7 mr beam divergence	70
2. Ruby	0.6943	140	0.5 msec	12 mm x 15 cm		71

*Pulse repetition frequencies are typically one pulse per minute.

Table 5-10. Typical high energy laser oscillators.*

Pulse durations of a few milliseconds are usual. Peak powers may be several hundreds of kilowatts.

The energy conversion efficiency is about 1.0 percent for ruby and as high as 6.0 percent for glass. Spectral linewidths vary. For multimode glass it could be as high as 100Å - much less for ruby.

The beam divergence of high energy laser oscillators is typically of the order of 100 milliradians.

Pulsed Laser Oscillators - High Peak Power. High peak powers are obtained by using the "Q-switched" mode of operation. This is an energy storage mode rather than a conversion mode. The energy is stored in the laser rod and released impulsively. The output is a smooth, very short pulse of monochromatic radiation which is contained in a very small solid angle.

The operating characteristics of a few of these devices are summarized in Table 5-11.

Peak powers as high as 700 megawatts from a single ruby oscillator have been achieved. The pulse duration is usually of the order of a few nanoseconds with pulse energies of a few joules being quite typical.

Material	Wavelength (microns)	Peak Power (megawatts)	Pulse Energy (Joules)	Rod Size	Comments	Reference
1. Ruby	0.6943	50	1.0	---	Commercially available, beam divergence = 7 mr	72
2. Ruby	0.6943	80	2.8	10 mm x 7.6 cm	beam divergence = 0.2 mr	73
3. Ruby	0.6943	500	4.0	---	Commercially available; beam divergence = 7 mr	70
4. Ruby	0.6943	700	--	13 mm x 15.2 cm		74
5. Nd:glass	1.06	30	1.5	6 mm x 46 cm		75
6. Nd:glass	1.06	100	--	---	U. of Moscow	74

*Pulse repetition frequencies are typically one pulse per minute.

Table 5-11. High peak power laser oscillators.*

The energy conversion efficiency of this type of operation is lower than in the high energy mode - typically about 0.1 percent. Spectral linewidths can vary from 0.001Å for single mode operation in ruby to 100Å for multimode operation in Nd:glass. Correspondingly, the beam divergence for single mode operation can be as small as approximately 0.2 milliradians and for multimode operation may be as large as approximately 10 milliradians.

At the present time, these devices can operate at a pulse repetition frequency of about one pulse every few seconds. Cooling of the laser components is the major problem to be overcome in achieving higher repetition frequencies.

Pulsed Laser Oscillators - High Pulse Repetition Frequency (PRF). Pulsed laser oscillators with a PRF of 100 pps or higher have been considered for discussion in this section. Table 5-12 contains information on a number of these devices. Pulse repetition frequencies as high as 12 kcps have been achieved in a He-Ne laser and 5 kcps in Nd:YAG. Peak powers vary from a few watts at very high rep rates to several megawatts with Q-switched operation at lower rep rates.

Material	Wavelength (Microns)	Peak Power	Pulse Energy	Repetition Rate	Comments	Reference
He-Ne	1.118	300 w	0.25 mj	2.6 kcps	ave. power = 0.6 w	76
Ar^+	several lines 0.4579-0.5145	50 w	0.05 mj	200 cps	ave. power = 0.1 w	77
Nd:YAG	1.06	250 w	0.05 mj	100 cps	cw pump, output power not optimized	78
Nd:YAG*	1.06	2 kw	0.2 mj	5 kcps		
Nd:$CaWO_4$	1.06	6 mw	140 mj	100 cps	ave. power 14 w	79
GaAs	0.84	2 w	2 μj	10 kcps	77°K, 7 Å, efficiency = 1%	80
GaAs	0.90	4 w	0.2 μj	1 kcps	300°K, 14 Å, efficiency = 0.04%	81

*Goal of a present effort at HAC according to private communication from E. J. Woodbury.

Table 5-12. High repetition frequency laser oscillators.

Pulse energies are quite small, usually of the order of a millijoule, with the result that the average power ranges from a few milliwatts to a maximum of about 15 watts.

When the repetition rate is comparable to build up time the efficiency of a high prf system can be severely degraded if repetition is secured by modulating the pump. Conversely if the prf is comparable to relaxation time, no loss in efficiency is obtained by using cw pumping and modulating the regeneration cavity, either by Q modulation or other intracavity modulation techniques.

Solar Pumped Lasers. Some investigators have succeeded in pumping lasers directly with sunlight. This would eliminate some of the power supply problem in regions where the ambient solar flux was high enough, i.e., Mercury to Mars missions. The only significant problem would be the necessity of continuously pointing the solar energy collector at the sun and folding the laser output path so as to continuously point to earth.

G.R. Simpson of American Optical Company has discussed[82] a solar pumped Nd_2O_3 - doped barium crown glass. He obtained continuous oscillation at 1.06 μ. The laser rod was 30 mm long by 1 mm diameter. The solar energy collector was a 61-cm diameter, f/1.5 parabolic mirror. The arrangement is shown in Figure 5-41.

P.H. Keck, et al of Texas Instruments Inc., have described[83] a proposed optical condenser to effectively concentrate sunlight from a parabolic collector onto a laser rod. This is shown in Figure 5-42. A solid cone-sphere condenser of Sr Ti O_3 was tested using a xenon arc lamp source. A 12 mm long by 1 to 2 mm diameter laser rod of neodymium-doped calcium tungstate was pumped continuously. The arrangement is shown in Figure 5-43.

Hughes Aircraft Company has done developmental work on lightweight solar concentrators[84] which may have possible application to direct solar pumping of lasers.

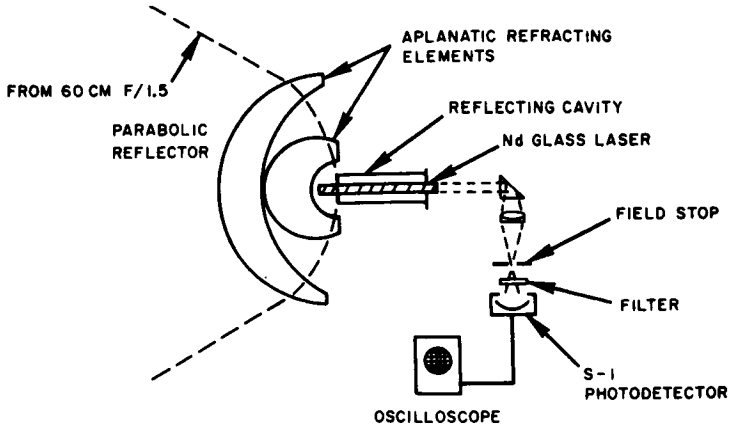


Figure 5-41. Schematic illustration of pumping and detection scheme for sun-pumped laser.

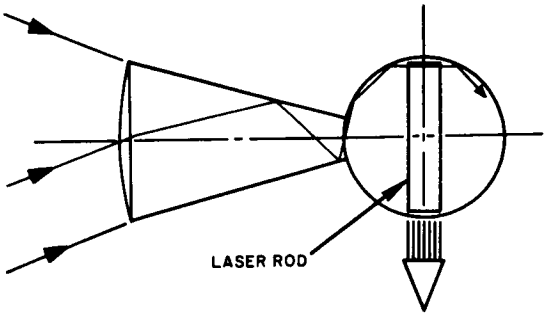


Figure 5-42. A cone-sphere condenser.

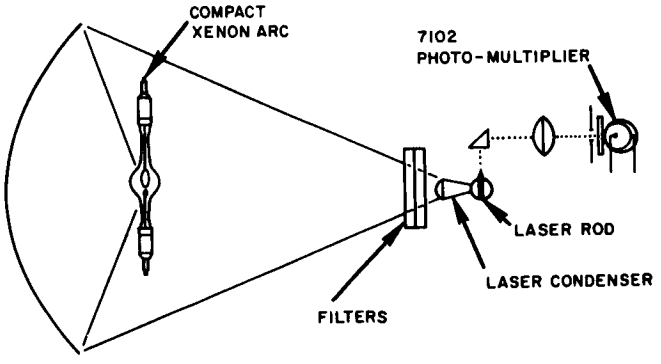


Figure 5-43. Experimental optical setup for pumping laser rod and detecting laser output.

Laser Amplifiers. A more elaborate laser transmitter may employ one or more laser amplifiers to increase the output power and energy. Another use of laser amplifiers is to amplify single mode oscillators and thus have high spectral radiance.

Laser amplifiers may be operated on a cw-pump or pulse basis. The operating characteristics of interest are:

1. small signal power gain
2. saturation power
3. small signal energy gain
4. saturation energy
5. bandwidth
6. operating wavelength
7. distortion
8. noise
9. efficiency

Cw Laser Amplifiers. At the present time, only gas lasers are used as cw laser amplifiers, because cw amplifiers are generally low power devices and only gas devices have found application in the low power cw regime.

The small signal power gain of a gas laser can be quite high (as much as 70 db/m for xenon); however, the saturation power density at the output aperture is usually of the order of a few milliwatts per cm^2 so high power operation is not generally practical.

The 3 db bandwidth of a laser amplifier is related to the amplifier gain and the fluorescent linewidth of the laser transition by (for large gain)

$$(3 \text{ db bandwidth}) \cong (\text{fluor. linewidth}) \; / \sqrt{\ell n G} \; .$$

Fluorescent linewidths vary from 100 mc/s to 5,000 mc/s for gas lasers. Thus, for a gain of 20 db, the 3 db bandwidth of the laser amplifier will be narrowed by a factor of approximately two.

The relatively narrow passband of a laser amplifier requires that the operating wavelength of the amplifier and oscillator be closely matched. The usual way to accomplish this is to use the same laser material for both oscillator and amplifier. Even in this case, however, a temperature differential may be sufficient to put the oscillator frequency outside the passband of the amplifier. The oscillator-amplifier frequency matching problem can be alleviated somewhat by the use of laser frequency tuning techniques. These allow one to change the frequency by small amounts, generally by less than one part in 10^4. Nonlinear devices may also be used to achieve frequency diversification; but have not been treated in this report. The distortion or spreading of the laser beam in passing through an amplifier depends on the optical homogeneity of the laser amplifier medium. For gas laser amplifiers this is not a problem. A diffraction limited input beam yields a diffraction limited output beam.

Amplifier noise is due to spontaneous emission. It increases with the gain of the amplifier according to the relation

$$P_s = h\nu \, \Delta f(G-1)$$

where $h\nu$ is the photo energy, Δf is the passband of the amplifier, and G is the gain. At 10 db gain, gas laser amplifiers produce noise powers ranging typically from 10^{-10} to 10^{-8} watts.

Pulsed Laser Amplifiers. For pulsed laser amplifiers the input pulse power is always sufficient to cause power gain saturation. For very short pulses, only the leading edge of the input pulse sees the gain of the fully pumped amplifier. Subsequent portions of the input pulse see a smaller gain due to depopulation by the leading edge of the pulse. The result of this saturation phenomena is a "pulse-shaping" or distortion of the input pulse envelope. Small signal power gains of 30 db in a 15-cm ruby rod have been achieved. Due to the finite rise time of the input pulse generated by a laser oscillator, the peak power gain is more like 10 to 15 db for the same rod.

Since power gain varies over the duration of the pulse, it is sometimes convenient to define an energy gain as (output pulse energy)/(input pulse energy).

Small signal energy gains of 20 db in a 15-cm ruby rod have been achieved. Saturation occurs when the input pulse energy density (joules/cm^2) is sufficient to completely depopulate portions of the amplifier rod. For ruby and Nd:glass this occurs at approximately 7 joules/cm^2; for Nd:YAG at 0.06 j/cm^2. Solid state materials are used for high power or high energy laser amplifiers due to their energy storage capability. Ruby and Nd:glass can store typically from 1 to 5 joules per cc. Consequently, the available wavelengths are 0.6943 microns in (ruby) and 1.06 microns in (Nd^{+++}). Pulsed solid state laser amplifiers have larger passbands than cw gas laser amplifiers; the passband for ruby is about 1 Å and for Nd:glass, it is several tens of Å.

Distortion in pulsed amplifiers in the form of increased beam divergency may exist due to transient thermal gradients in the amplifier material. However, operation of ruby amplifiers with no spreading of the input beam in passing through the amplifier has been demonstrated.

Two factors tend to limit oscillator-amplifier performance. These are damage thresholds for power density and for energy density. If the power density in a dielectric becomes too high, the resultant electric field can cause dielectric breakdown of the material. If the energy density in a pulse, whose duration is short compared to some thermal relaxation time, becomes large, energy will be absorbed by the material faster than it can be carried away and the material will melt. This seems to be the most important factor limiting the performance of high energy glass laser amplifiers. The efficiency of ruby and Nd:glass laser amplifiers approximates the values obtained in high energy quasi-cw conversion mode oscillators.

5.3.7 Power/Burden Relationships

Establishment of Measure of Merit. In making a relative comparison of transmitter sources at various wavelengths, it is best to consider first a general figure of merit which ignores many of the practical limitations. In this way, the considerations are not arbitrarily bounded or complicated too early in an evaluation. Also, as the practical limitations are included, their effects can be evaluated individually. In this way, areas where substantial research or development may be beneficial can be more readily identified.

In considering gas laser sources, an appropriate figure of merit is

$$M_L = \frac{\eta^2 P}{L\lambda^2}$$

where L is the length of the laser. This is useful since most lasers are rated in power per unit length. An appropriate weight per unit length must then be determined in order to make direct comparisons with microwave sources.

For ground based transmitters, it is believed that only the maximum power available at a given wavelength and the antenna gain are significant. Accordingly, the figure of merit for ground based sources is simply

$$M_G = P_{max}/\lambda^2$$

These figures of merit will be discussed in more detail and refined as the description of various laser and microwave sources progresses in subsequent editions of this report.

The following general assumptions have been made in choosing to compare lasers on the basis of P/λ^2 for ground based transmitters and $\eta^2 P/L\lambda^2$ for spacecraft transmitters:

1. No atmosphere attenuation at the laser wavelengths.
2. No prime power limitations for ground transmitter. Penalty of η^2 for the spacecraft transmitter.
3. No length penalty for ground transmitter; penalty of $1/L$ for spacecraft.
4. No limitation due to mirror damage on ground transmitter.
5. No pointing accuracy limitations on either transmitter.
6. No optics limits.
7. No cost limitations.
8. Spectral output suitable for the communications job required.

Gas Laser Sources Evaluation. Table 5-13 and Figures 5-44 and 5-45 summarize the characteristics of the six best gas laser sources reported to date using the above figures of merit. Hundreds of other wavelengths are available, but these six have been selected as the best and most representative of each type (ion, molecular, and neutral gas). The reported output power, length and input power are given for the lasers selected. From these, the parameters P/λ^2 and $\eta^2 P/L\lambda^2$ are computed. For comparison with microwave devices, where the parameter $\eta^2 P/W\lambda^2$ has been used, we may assume a value of W/L of about 20 lbs/meter for He-Ne, CO_2 and He-Xe lasers and about 50 lbs/meter for ArII. The exact W/L will depend on L and the stability requirements for the particular application. However, these are reasonably realistic numbers for the sake of comparison. The heavier value for ArII reflects the additional requirement of a magnetic field (~ 1,000 gauss).

Notes

1. This is a Hughes airborne quartz laser with a 46 cm bore length, ~ 1 meter overall package length. It requires a magnetic field of ~1,000 gauss, which would imply a heavy structure and possibly more power.

Gas	λ (μ)	Manufacture	P out (W)	L (M)	P in (W)	η	P/λ^2 (W/M^2)	$\eta^2 P/L\lambda^2$ (W/M^3)	Note
Ar II	0.5	HRL	4.0	0.46	4,000	1×10^{-3}	1.6×10^{13}	3.5×10^{7}	1
Ar II	0.5	RAY	8.0	1.6	20,000	4×10^{-4}	3.2×10^{13}	3.2×10^{6}	2
He-Ne	0.63	BTL	1.0	5	500	2×10^{-3}	2.5×10^{12}	2.0×10^{6}	3
He-Ne	0.63	S-P	0.1	1.7	~200	5×10^{-4}	2.5×10^{11}	3.7×10^{4}	4
He-Ne	1.15	S-P	0.03	1.7	~200	1.5×10^{-4}	2.3×10^{10}	3.0×10^{2}	5
He-Ne	3.39	HRL	0.01	1.7	80	1.8×10^{-4}	8.7×10^{8}	1.7×10^{1}	6
He-Xe	3.51	TRG	0.08	2.0	~200	4×10^{-4}	6.5×10^{9}	5.2×10^{2}	7
CO_2	10.6	HRL	10	2.0	150	6.7×10^{-2}	8.9×10^{10}	2.0×10^{8}	8
	10.6	BTL	12	2.0	-	3×10^{-2}	1.1×10^{11}	4.8×10^{7}	9
	10.6	BTL	130	4.0	~1,000	$\sim 1.3 \times 10^{-1}$	1.2×10^{12}	4.9×10^{9}	10
	10.6	BTL	0.1	0.5	30	3.3×10^{-3}	8.9×10^{8}	1.9×10^{4}	11

Table 5-13. Figure of merit for several gas lasers.

2. This laser was reported by Raytheon in Electronic News; It is a quartz tube. 18 watts was claimed, provided the beam in the cavity was chopped to prevent damage to the mirrors. Figures were never officially published.
3. This is Bell Labs best effort in red lasers. It has never been reproduced elsewhere and requires "Bell Labs Mirrors." It is doubtful that anyone will ever attempt to build a more powerful 6328A laser.
4. This is a commercially available Spectra-Physics model 125. 50 mw is guaranteed, but selected tubes produce 100 mw. This is roughly the best achievable at 6328A in a laser this size.
5. This is the Spectra-Physics model 125. With more work it may be possible to double the power in a tube this size, but drastic improvements are quite unlikely at this wavelength.

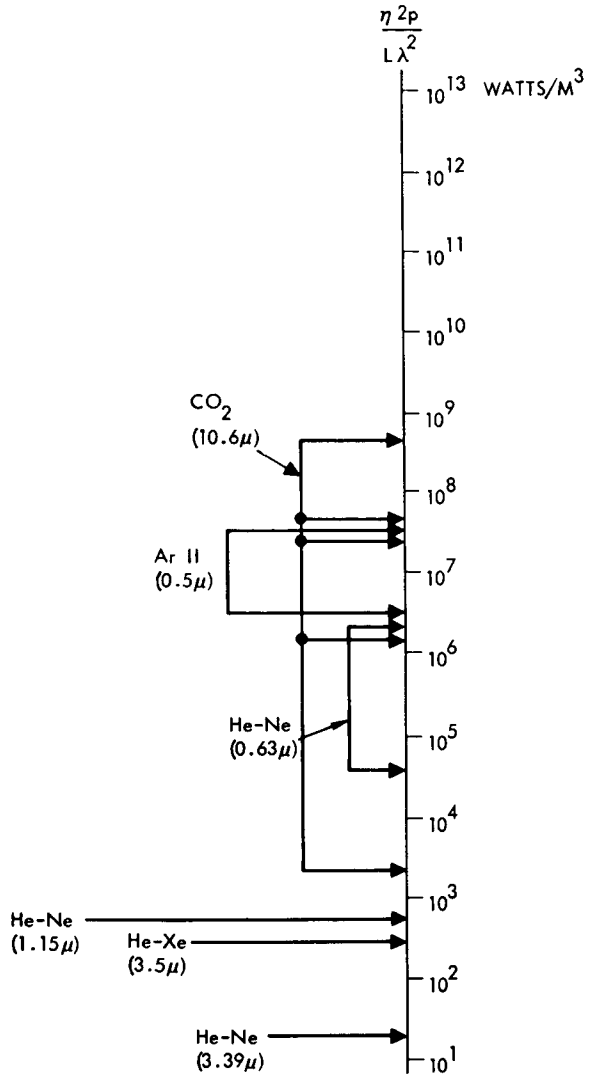


Figure 5-44. Comparison of lasers for space-craft transmitter.

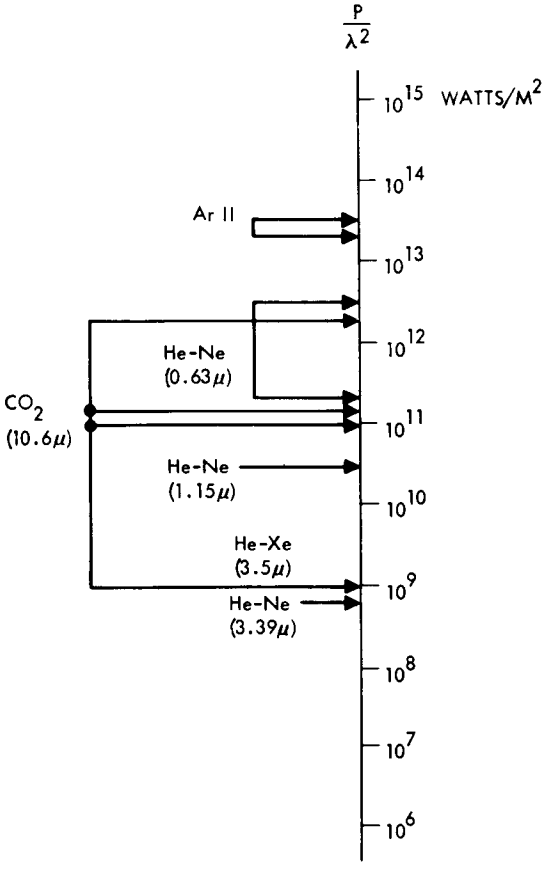


Figure 5-45. Comparison of lasers for ground based transmitter.

6. This is an HRL Lab-type tube. With more work the output may be doubled, but more than this is doubtful in a tube this size.
7. This is a TRG Laboratory type tube and represents approximately two years of effort in developing a high power Xe laser. It is probably close to the ultimate for a tube this size.
8. This is an HRL Laboratory type tube using flowing CO_2-N_2 mirrors were not optimized; more output power can be

expected from this same tube (~20 watts). Seven watts were obtained with the tube sealed off.

9. This is the best published BTL work (C.K.N. PATEL, Appl. Phys. Lett., 1 July 1965). A flowing gas system was used with a mixture of CO_2, N_2, O_2, H_2O.
10. This is unpublished BTL result with a tube 4 inches in diameter and 12 feet long. Helium was used. It is hard to estimate how much will eventually be obtained from this tube but an intensive effort has been underway for several months. (C.K.N. Patel).
11. This is a small, non flow tube, with external mirrors; suitable for spacecraft. This work is due to T. J. Bridges (unpublished) and is rather preliminary. A recent account of similar tubes appears in the 1 November 1965 Appl. Phys. Letters.

In comparing the various lasers listed in Table 5-13, the suitability of the output signal for the communications task at hand must be kept in mind. All of the lasers listed can be made to operate in the lowest order spatial mode (TEM_{oo}) alone with more or less difficulty. The task is easier at the shorter wavelengths where the laser output is visible and the characteristic beam size is small (proportional to the square root of the product of wavelength and a cavity parameter related to mirror radius). Mode selection at infrared wavelengths may be done with an image-converter or by listening to self-beats in a heterodyne detector. Production of a single-frequency output is still quite difficult because of the longitudinal mode structure of the long Fabry-Perot cavities used. Only the 10.6μ CO_2 and 3.5μ Xenon lines are narrow enough to produce reasonable output by keeping the Fabry-Perot resonator short enough so that only one longitudinal mode oscillator. This is done to the narrow doppler-broadened line widths of these two transitions (≈ 50 mc for 10.6μ CO_2 and ≈120 mc for 3.5μ Xe). Even these two transitions will require further mode selection techniques if longer, higher power tubes are considered. Because of the broad doppler line widths of the Ar and He-Ne lasers, single-frequency operation through

the use of a sufficiently short Fabry-Perot resonator entails a drastic loss in output power. Techniques involving 3 mirror resonators allow the use of longer tubes at the expense of added complexity both mechanical and electronic (servo-controlled mirror positioning), but still sacrifice output power because the entire line is not used. The most promising technique developed to date is that of intracavity mode locking[11] with a subsequent coherent recombination[85] or selective output coupling[86]. This technique has been demonstrated in the laboratory, but practical power levels at a single frequency are yet to be obtained. In any case the additional complexity will contribute to the weight, length and inefficiency of the laser, although perhaps not to a significant extent.

It appears that, at present, the best* combination for optical space communications from Figure 5-44 and 5-45 would be a small, efficient, light weight 10.6μ CO_2 laser in the spacecraft with coherent detection (superheterodyne) on the ground, employing a cooled Hg:Ge detector. The up-link would be best handled by a high-power multimode argon laser on the ground, employing pulse amplitude or pulse polarization modulation, and a simple ruggedized photomultiplier video receiver in the spacecraft. These conclusions are of course, subject to revision as the state of the laser (and detector) art progresses.

5.4 NOMENCLATURE

The following list of nomenclature is used in section 5.0.

A	- junction area
A_2	- net total spontaneous emission coefficient for lower laser level
A_3	- net total spontaneous emission coefficient for upper laser level
B	- magnetic field strength

*Strictly from the transmitter viewpoint, of course.

$B'_{23}(\nu', \nu)$; $B'_{32}(\nu', \nu)$ - simulated emission rates

b_1, b_2 - mirror radii

C - capacitance of depletion avalanche zones

c - speed of light

D - diameter

E - average longitudinal electric field

f - frequency

G - amplifier gain

go - gain coefficient (meter^{-1})

I; i - current

I_o - bias current

J - current density

L; ℓ - lengths

L_a, L_b - length of interferometer paths

ℓ_o - length of depletion zone

M_{SB}; M_G; M_A - figures of merit

m, n, q - mode numbers

N_1 - number density in lower level

N_2 - number density in upper level

n - index of refraction

n_2 - volume densities of atoms in lower level

n_3 - volume densities of atoms in upper level

P - power

P_s - amplifier noise

R_s - bulk resistance of region outside depletion zone

S - power circulating in cavity

S_{3db} - power level

T - temperature

V_{cath} - Cathode fall

V_P - Saturation velocity

W - weight

Z - equivalent impedance of diode

z - length of tube

α - ionization coefficient

α' - derivative of α with respect to electric field

Γ - modulation index

Γ' - modulation index of external modulator

Δf - passband of amplifier

$\Delta \nu$ - spontaneous emission linewidth; Fabry-Perot transmission bandwidth (3 db)

$\Delta \nu'$ - frequency difference between $\Delta\Omega$ and ν_m

$\Delta \nu_D$ - Doppler linewidth (3 db)

$\Delta \nu_N$ - natural linewidth (3 db)

$\Delta\Omega$ - frequency separation of modes

δ - phase retardation (radians)

δL - change in cavity length

δn - change in index of refraction

δp - change in pressure

δT - change in temperature

$\delta \nu$ - change in optical frequency

η - efficiency

θ - transient time through the depletion layer

λ - wavelength

$\dot{\mu}$ - thermal coefficient of expansion

ν - optical frequency

ν_o - center frequency

ν_1, ν_2, ν_3 - particular optical frequencies

ν_m - modulating frequency

ρ - doping density

ϕ - difference in phase of two modulators

ω_a - avalanche frequency

5.5 REFERENCES

1. Matthei, W. G., "Recent Developments in Solid State Microwave Devices," Microwave Journal, Vol. 9, No. 3, P. 39, March 1966.

2. Gilden, M. and Hines, M. E., "Electronic Tuning Effects in Read Microwave Avalanche Diodes," IEEE Trans on Electron Devices, Vol. ED-13, No. 1, p. 169, January 1966.

3. Brand, F. A., et al, "Performance Characteristics of CW Silicon and GaAs Avalanche Diode Oscillators," G-MTT Symposium, May 16-19, 1966, Palo Alto, California.

4. Gunn, J. B., "Instabilities of Current in III-V Semiconductors," IBM Journal Res. Develop., Vol. 18, No. 2, April 1964, pp. 141-159.

5. Thim, H. W., et al, "Microwave Amplification in a dc Biased Bulk Semiconductor," Appl. Phys. Letters, Vol. 7, p. 167, September 1965.

6. Hakki, B. W., and Knight, S., "Microwave Phenomena in Bulk GaAs," IEEE Transactions on Electron Devices, Vol. ED-13, No. 1, p. 94, January 1966.

7. A. Yariv, J. Appl. Phys., 36 388-391, Feb 1965.

8. M. DiDomenico, Jr., J. Appl. Phys., 35 2870-2876, Oct 1964.

9. M. H. Crowell, IEEE J. of Quantum Electronics, QE-1, 12-20, April 1965.

10. L. E. Hargrove, R. L. Fork, and M. A. Pollack, Appl. Phys. Letters, 5, 4-5, 1 July 1964.

11. S. E. Harris and O. P. McDuff., Appl. Phys. Letters 5, 205-206, Nov. 15, 1964.

12. S. E. Harris and O. P. McDuff, IEEE journ. of Quantum Electronics QE-1, 245-262, Sept. 1965.

13. E. O. Amman, B. J. McMurtry, and M. K. Oshman, IEEE J. of Quantum Electronics, QE-1, 263-272, September 1965.

14. G. A. Massey, M. K. Oshman, and Russell Targ, Appl. Physl Letters 6, 10-11, 1 Jan 1965.

15. S. E. Harris and B. J. McMurtry, Appl. Phys. Letters, 7, 265-267, 15 Nov. 1965.

16. R. Targ and B. J. McMurtry, Paper presented at the 1966 International Quantum Electronics Conference, Phoenix, Arizona April 12-15, 1966. To be published.

17. L. Osterink, R. Byers, and S. E. Harris, Paper presented at the 1966 International Quantum Electronics Conference., Phoenix Arizona, April 12-15, 1966. To be published.

18. A. D. White, "Frequency stabilization of gas lasers," IEEE J. Quantum Electronics, QE-1 pp 349-357, Nov. 1965.

19. G. D. Boyd and H. Kogelnik, "Generalized confocal resonator theory", Bell Sys. Tech. J. 41, pp. 1347-1370, July 1962.

20. I. P. Kaminow, "Balanced optical discriminator," Appl. Optics. 3, pp. 507-510, April 1964.

21. S. E. Harris, "Demodulation of phase-modulated light using biretringent crystals", Proc. IEEE, 52, pp. 823-831, July 1964.

22. W. R. Bennett, Jr., "Mode pulling in gas lasers," in Quantum Electronics III, P. Grivet and N. Bloembergen, Eds. New York: Columbia University Press, 1964.

23. A. D. White, E. I. Gordon, and E. F. Labuda, "Frequency Stabilization of single mode gas lasers", Appl. Phys. Letters, 5, p. 97-98, September 1964.

24. I. Tobias, M. Skolnik, R. A. Wallace, and T. Polanyi, "Derivation of a frequency-sensitive signal from a gas laser in an axial magnetic field", Appl. Phys. Letters, 6, pp. 198-200, May 1965.

25. W. R. Bennett, Jr., S. F. Jacobs, J. T. Latourette, and P. Rabinowitz, "Dispersion characteristics and frequency stabilization of a gas laser", Appl. Phys. letters, 5, pp. 56-58, August 1964.

26. J. A. Collinson, "A Stable, Single-Frequency r-f excited gas laser at 6328A", Bell Sys. Tech. J. 44, pp 1511-1519, Sept. 1965.

27. W. R. C. Rowley and D. C. Wilson, "Wavelength stabilization of an optical maser", Nature, p. 745, November 1963.

28. K. Shimoda and A. Javan, "Stabilization of the He-Ne maser on the atomic line center", J. Appl. Phys., 36 pp. 718 - , March 1965 (pt. 1).

29. A. D. White "A two-channel laser frequency control system", IEEE J. Quantum Electronics, QE-1, pp 322-333, October 1965.

30. C. K. N. Patel, W. L. Faust, and R. A. McFarlane, "Laser action on rotational transitions of the $\Sigma_u^+ - \Sigma_g^+$ vibrational band of CO_2," Bull. Am. Phys. Soc. 9, 500 (1964).

31. C. K. N. Patel, "interpretation of CO_2 optical maser experiments," Phys. Rev. Letters 12, 588 (1964).

32. C. K. N. Patel, "Selective excitation through vibrational energy transfer and optical maser action in N_2-CO_2," Phys. Rev. Letters 13, 617 (1964).

33. N. Legay-Sommaire, L. Henry, and F. Legay, "Réalisation d'un laser utilisant l'énergie de vibration des gaz excités par l'azote activé (CO, CO_2 et N_2O)," Compt. Rend. 260, 3339 (1965).

34. P. Barchewitz, L. Dorlec, A. Truffert, and P. Vautier, "Émission laser continue par excitation haute fréquence directe de CO_2 et N_2O dans les transitions vibrationelles $00^01 - 10^00$ et $00^01 - 02^00$," Compt. Rend. 260, 5491 (1965).

35. C. K. N. Patel, "Vibration energy transfer, and efficient means of selective excitation," Phys. of Quant. Elect. Conf., San Juan, Puerto Rico, June 29-30, 1965.

36. W. J. Witteman, "increasing continuous laser action on CO_2 rotational-vibrational transitions through selective depopulation of the lower laser level by means of water vapor," Phys. Letters 18, 125 (1965).

37. R. A. McFarlane and J. A. Howe, "Stimulated emission in the system CO/CO_2," Phys. Letters 19, 208 (1965).

38. T. J. Bridges and C. K. N. Patel, "high-power Brewster window laser at 10.6 microns," Appl. Phys. Letters 7, 244 (1965).

39. G. Moeller and J. D. Rigden, "high-power laser action in CO_2-HE Mixtures," Appl. Phys. Letters 7, 274 (1965).

40. C. K. N. Patel, P. K. Tien, and J. H. McFee, "Cw high-power CO_2-N_2-He laser," Appl. Phys. Letters 7, 290 (1965).

41. C. K. N. Patel, "Continuous-wave laser action on vibrational-rotational transitions of CO_2," Phys. Rev. 136, A1187 (1964).

42. P. Barchewitz, L. Dorbec, R. Farrenq, A. Truffert, and P. Vautier, "Émission infrarouge de CO et CO_2 et laser continu a CO_2 par action directe d'une excitation haute fréquence," Compt. Rend. 260, 3581 (1965).

43. C. K. N. Patel, "Cw high power N_2-CO_2 laser," Appl. Phys. Letters 7, 15 (1965).

44. R. Farrenq, C. Meyer, C. Rossetti, L. Dorbec, and P. Barchewitz, "Amplification du rayonnement d'un laser à CO_2 par du gaz carbonique excite en haute frequence," Compt. Rend. 261, 2617 (1965).

45. J. A. Howe, "Effect of foreign gases on the CO_2 laser: R-branch transitions," Appl. Phys. Letters 7, 21 (1965).

46. G. Moeller and J. D. Rigden, "Observations of laser action in the R-branch of CO_2 and N_2O vibrational spectra," Appl. Phys. Letters 8, 69, (1966).

47. C. Frapard, "Vibrational excitation of CO_2 by electronic impact in pure CO_2 laser" presented at 1966 International Quantum Electronics Conference, Phoenix, Arizona, April 1966.

48. J. A. Howe and R. A. McFarlane, "New emission systems in CO_2", J. of Mol. Spect. 19, 224, (1966).

49. H. Statz, C. L. Tang and G. F. Koster, "Probabilities for radiative transitions in the CO_2 laser system," presented at 1966 International Quantum Electronics Conference, Phoenix, Arizona, April 1966.

50. R. N. Schwartz, Z. I. Slawsky, and K. F. Herzfeld, "Calculation of Vibrational Relaxation Times in Gases," J. Chem. Phys., Vol. 20, pp. 1591-1599, (October 1952).

51. R. Marriott, "Molecular Collision Cross-Sections and Vibrational Relaxation in Carbon Dioxide," Proc. Phys. Soc., Vol. 84, pp. 877-888 (1965).

52. R. Marriott, "Molecular Collision Cross-Sections and the Effect of Hydrogen on Vibrational Relaxation in Carbon Dioxide," Proc. Phys. Soc., Vol. 86, pp. 1041-1049 (1965).

53. D. R. Bates, "Collision processes not Involving Chemical Reactions," Discussions Faraday Soc. 33, 7 (1962).

54. H. J. Bauer and E. Liska, "Schwingungsrelaxation von CO_2 in Gemischen mit Helium," Zeit. f. Physik, Vol. 181, pp. 356-367 (December 1964).

55. A Javan and P. L. Kelley, "Possibility of self-focusing due to non-linear anomalous dispersion," Presented at the 1966 International Quantum Electronics Conference, Phoenix, Arizona, April 1966.

56. M. A. Kovacs, G. W. Flynn, and A. Javan, "Q-switching of molecular laser transitions," Appl. Phys. Letters 8, 62, 1966.

57. Spectra-Physics Model 125 Gas Laser, Spectra-Physics, Inc., Mountain View, California.

58. D. L. Perry, "Mirror Coating Procedures for High Power Gas Lasers," 1964 NEREM (November) paper, p. - 3.5.

59. Hughes Research Laboratories Final Report, JPL Contract No. 950803 (April 1965).

60. E. I. Gordon, private communication

61. R. Paananen, "Progress in High Power Ionized Argon Lasers," Proceedings of the Second Conference on Laser Technology, Chicago, Ill., 6-8, April 1965.

62. Hughes Research Laboratories Model 50 Argon Laser W. B. Bridges, private communication

63. C. K. N. Patel, Bell Laboratories Record, July/August 1965

64. P. K. Tien, Bell Laboratories, private communication

65. V. Evtuhov, Appl. Phys. Left. 6, 75 (1965)

66. H. R. Aldag, Appl. Optics 4, 559 (1965)

67. J. E. Geusic, Appl. Phys. Lett. 6, 175 (1965)

68. KORAD Model KY-1; Korad Corp., Santa Monica, California

69. RCA Interim Engineering Report, No. 2, Solid State Laser Explorations; May 1965

70. KORAD Model K-2Q; Korad Corp., Santa Monica, California

71. KORAD Model K2; Korad Corp., Santa Monica, California

72. KORAD Model K-1Q; Korad Corp., Santa Monica, California

73. E. R. Peressini, Appl. Phys. Lett. 3, 203 (1963)

74. W. R. Sooy, private communication

75. E. Snitzer, "Neothymium Glass Laser," Proceedings of the Third Int'l. Conf. on Quantum Electronics, Paris, France, 11-15 Feb. 1963

76. J. Goldsmith, "Measurement of High Power Output from a He-Ne Pulsed Gas Laser Employing an Exit Mirror of Optimum Reflectivity," MIL-E-CON Conference, Washington, D. C., 14-16 Sept. 1964.

77. W. B. Bridges, Private communication

78. C. G. Young, Appl. Phys. Lett, 2, 151 (1963)

79. A. Gilmer, private communication

80. KORAD Model KS-3; Korad Corp., Santa Monica, Calif.

81. KORAD Model KR-2; Korad Corp., Santa Monica, California

82. Simpson, G. R., Applied Optics, 3, 6, pp 783, 4, June 1964

83. Keck, P. H., et al Applied Optics, 2, 8, pp 827 to 831, Aug. 1963

84 Bagley, J. P. and Schwartz, S., "Rigidized Inflatable Solar Energy Concentrators," NASA CR-254, Washington, D. C. (July 1965)

85. Massey, G. A., M. K. Ashman and R. Taig, APL 6 10 (1965)

86. S. E. Hanes and B. J. McMurtry (to be published)

7.0 OPTICAL MODULATORS

In the present revision, being made approximately four months after preparation of the first edition, emphasis is placed on the more important advances in optical modulation technology which are currently in progress. A section on Elasto-Optic Modulation has been added. This is of timely importance because of the urgent need to develop techniques for modulating the high-power sources being made available at 10.6 microns by the CO_2 laser. The subject of internal modulation of laser oscillators has been separated from another part of the text and considerably expanded. This technology, similarly to that of acoustic (elasto-optic) modulation, is still in an early stage of development. Its importance lies in its promise to effect great economies in size, weight, and driving power required as compared with existing optical modulation devices which are external to the laser oscillator.

The section on Electro-Optic Modulation has been revised in several ways to improve and clarify the presentation. Data has been included on lithium niobate ($LiNbO_3$), which has recently become available with sufficient size and optical quality to show interesting promise at least for internal modulation. This crystal has good optical transmission from about 4000 Å to beyond 4 microns. Available data on lithium tantalate, another promising new crystal, also has been added.

The possibility of placing some initial approximate values on the modulator performance/burden relationships defined in Section 2 was considered. In our judgment it is still premature to make meaningful assessments of the weight, cost, or driving power required as a function of information bandwidth. These values will vary with the particular optical source considered, and may not be continuously variable as a function of information bandwidth. At 10.6 microns, a wavelength of great current interest for space applications, there is essentially no existing technology for wideband modulation of high-power

optical beams. A speculative approach is to use acoustic techniques applied internally to the laser oscillator. There is, however, no valid basis at this time for making a prediction of the performance/burden relationships which might result from development of this technique. The best hope that can be foreseen at this time is that such values for selected laser system designs will be determined at a later time during this project.

7.1 INTRODUCTION

Laser modulation is accomplished by passing the beam through an optically transmissive medium in which one or more of the optical transmission parameters is varied by the application of a modulating field. The interaction of the laser beam and the modulating field in the presence of an optically nonlinear medium makes it possible to achieve a wide variety of types of optical modulation, including intensity, frequency, phase, and polarization.

The advent of lasers has motivated extensive research and development in optical modulation. Rather primitive devices presently are available for obtaining all forms of modulation of visible lasers. Most of these devices are based on the use of the electro-optic effect in crystals. The best immediately available techniques for modulating infrared lasers (beyond about 1.5 microns) involve the use of elasto-optic, or acoustic effects. However recent advances in the technology of growing single-domain ferro-electric crystals, some of which exhibit a strong first-order electro-optic effect, promise a brighter future for modulation well beyond 1.5 microns by electro-optic means.

A detailed assessment of the performance and burdens of existing visible and infrared optical modulators may be misleading to the designer of future laser systems for space applications because the relevant technology is in a constant state of flux and probably will continue to be for at least several years. The discovery and perfection of nonlinear optical materials and the invention and perfection of modulation interaction structures and devices will lead to continuing improvement in performance and reduction of burdens.

At present, electro-optic modulators for lasers which operate in the wavelength range between about 0.4 and 1.5 microns can provide information bandwidths larger than the maximum believed to be required by space optical communication and tracking systems (i.e., 100 mc of bandwidth). The burdens of weight and modulator driving power, however, leave much to be desired. The problems associated with handling very high laser beam powers also have not yet been studied, and may present some additional difficulties. However, advances both in the synthesis of better electro-optic materials, and in the design of modulator structures will provide optical modulators suitable for space communications systems within the next few years. It is difficult to predict the time schedule for the availability of new electro-optic crystals. Notwithstanding such uncertainties, electro-optic materials capable of operating in the infrared region out to 10 microns have a good chance of becoming available.

In the meantime, it is necessary to intensify existing research in acoustic and other means for modulating infrared lasers above 1.5 microns. Present acoustic materials and techniques have serious limitations in achieving the larger information bandwidths that may be required (i.e., in extending present bandwidths of 1 to 5 mc, up to 100 mc).

Other physical processes which may be useful for optical modulation, such as controllable photon absorption in solid-state materials, and magneto-optics, either are not very promising, or are in too early a state of research to accurately evaluate their potential usefulness.

These introductory remarks on optical modulation have been addressed exclusively to the final modulation process of impressing the information on the optical carrier. It should be pointed out that there may be very good reasons for impressing the signal information on a radio-frequency subcarrier, prior to performing the final optical modulation. All of the conventional modulation techniques may be utilized in the preliminary step. This will eliminate the necessity for impressing video or very low-frequency modulation components on the laser beam. The elimination of low-frequency components on the optical

signal may provide important advantages both in the ease of design and driving of optical modulators, and in achieving satisfactory transmission through the earth's atmosphere. The latter factor is mentioned even though it is not yet certain that it will be either feasible or desirable to include the earth's atmosphere in the optical part of a space-earth link.

Finally, it should be pointed out that research and development remain to be done before the technology will exist to provide optical modulators with fully acceptable performance and burdens for space communication and tracking systems. The programs now in progress, if continued, may accomplish the desired results. This is in an area, however, which needs additional direction and support in order to assure that the results will be forthcoming on a suitable schedule.

7.2 ELECTRO-OPTIC MODULATION

7.2.1 Theory of Electro-Optic Modulation

Certain crystalline materials exhibit the property of induced birefringence when subjected to an electric field. If such a crystal is properly oriented between crossed polarizers in the path of a collimated monochromatic beam of light, an applied electric field will cause the incident plane-polarized beam to emerge elliptically polarized in general, and some of this light will thus pass through the second polarizer (analyzer). The axes of induced birefringence lie in a plane transverse to the beam direction and are oriented at ±45 degrees relative to the electric vector of the incident polarized light. It is shown in the following paragraphs that the intensity I of the beam passed by the analyzer, relative to the incident intensity I_o, is given by

$$I = I_o \sin^2 \frac{\Gamma}{2} . \qquad (7\text{-}1)$$

Γ is the relative optical phase retardation of the principal beam components along the fast and slow axes of induced birefringence, and is given explicitly by

$$\Gamma = \frac{2\pi \ell}{\lambda} (N_y - N_x)$$

where

ℓ ≡ length of crystal traversed by beam

λ ≡ optical wavelength

N_x ≡ refractive index of fast axis

N_y ≡ refractive index of slow axis.

Most currently available electro-optic crystals possess a first-order (Pockels) effect in which the induced birefringence is linearly proportional to the applied field. For this class

$$N_y - N_x = N_o^3 r E_m$$

where

N_o ≡ refractive index with zero applied field

r ≡ appropriate Pockels coefficient

E_m ≡ applied electric field.

Thus

$$\Gamma = \frac{2\pi \ell N_o^3 r E_m}{\lambda} \qquad (7\text{-}2)$$

A derivation of (7-1) follows.

Consider a vertically polarized light beam which traverses an electro-optic birefringent element followed by a horizontal polarizer (analyzer). The principal axes of birefringence x, y are oriented at 45 degrees relative to the electric vector of the incident polarized beam. At the front surface of the wave plate, the vertical and horizontal components of the beam may be represented by

$$\mathcal{E}_v = \mathcal{E} \cos \omega t, \qquad \mathcal{E}_h = 0 .$$

The principal components in the x, y frame are

$$\mathcal{E}_x = \frac{1}{\sqrt{2}} \mathcal{E} \cos \omega t, \qquad \mathcal{E}_y = \frac{1}{\sqrt{2}} \mathcal{E} \cos \omega t .$$

After traversing the wave plate, these components are out of phase by an amount Γ and, except for a common phase factor, are represented by

$$\mathcal{E}_x = \frac{1}{\sqrt{2}}\,\mathcal{E}\cos\left(\omega t + \frac{\Gamma}{2}\right), \qquad \mathcal{E}_y = \frac{1}{\sqrt{2}}\,\mathcal{E}\cos\left(\omega t - \frac{\Gamma}{2}\right),$$

assuming x to be the fast axis. The analyzer passes only the horizontal component of this beam. Thus the transmitted component is

$$\mathcal{E}_h = \frac{1}{2}\mathcal{E}\left\{\cos\left(\omega t + \frac{\Gamma}{2}\right) - \cos\left(\omega t - \frac{\Gamma}{2}\right)\right\} = -\,\mathcal{E}\sin\omega t\sin\frac{\Gamma}{2},$$

and its intensity I relative to that of the incident beam I_o is therefore

$$I = I_o \sin^2\frac{\Gamma}{2}.$$

In narrowband intensity modulation with sine wave rf drive, minimum harmonic distortion is achieved by biasing the incident beam with a quarter-wave plate to give circular polarization. In this case

$$\Gamma = \frac{\pi}{2} + \Gamma_o \sin\omega_m t.$$

We obtain

$$\frac{I}{I_o} = \frac{1}{2}\left[1 + \sin\left(\Gamma_o \sin\omega_m t\right)\right] = \frac{1}{2}\left[1 + 2J_1(\Gamma_o)\sin\omega_m t + 2J_3(\Gamma_o)\sin 3\omega_m t + \ldots\right].$$

When $\Gamma \le \frac{\pi}{6} \cong 0.5$, this can be written in the approximate form

$$\frac{I}{I_o} = \frac{1}{2}\left[1 + \Gamma_o \sin\omega_m t\right].$$

In this case $\Gamma_o \equiv m$, the usual designation of modulation index for intensity-modulated waves. More generally, however, there is a modulation index $m_i = 2J_i(\Gamma_o)$ for the fundamental and higher harmonics. For convenience, we list below some values of m_i for a selected range of Γ_o.

Γ_o	m_1	m_3
0.25	0.248	0.001
0.50	0.485	0.005
$\pi/4$	0.726	0.019
1.00	0.880	0.039
1.25	1.021	0.074
$\pi/2$	1.134	0.138

These figures show that if an attempt is made to achieve 100 percent modulation ($\Gamma = \pi/2$), the modulated wave contains almost 14 percent second-harmonic power. More significantly, since the driver power is proportional to Γ_o^2, it takes four times as much power to achieve 100 percent modulation as it does to achieve about 73 percent modulation ($\Gamma_o = \pi/4$), and nine times as much as that for 50 percent modulation ($\Gamma \cong \pi/6$). Note also that this last depth of modulation introduces only about 1/2 percent second-harmonic distortion.

The optical modulator used as an example for the present discussion employs KDP as the active electro-optic element. Table 7-1 shows the appropriate physical constants of this material. Reference numbers are shown in brackets.

The values of r_{63}, ϵ, and tan δ apply in the neighborhood of 150 mc. The Pockels constant is the "clamped" (zero-strain) value and is believed to be constant in the range of 0 to 500 mc.[1] Variation of the dielectric constant over this same frequency range is slight. The temperature coefficient of natural birefringence k is defined by the equation

$$\Delta\Gamma = \frac{2\pi\ell}{\lambda} k\Delta\Theta \qquad (7\text{-}3)$$

where $\Delta\Gamma$ represents the change in relative phase shift between ordinary and extraordinary rays in a length ℓ of crystal as a result of a change of temperature $\Delta\Theta$.[2]

Table 7-1. Optical and electrical constants of KDP.

Pockels constant r_{63}, m/V	8×10^{-12}	[7-1]
Ordinary refractive index N_o (at 6328 Å)	1.5073	[7-3]
Extraordinary refractive index N_e (at 6328 Å)	1.4670	[7-3]
Temperature coefficient of natural birefringence k, $^oC^{-1}$	1.1×10^{-5}	[7-2]
Dielectric constant ϵ, F/m	1.79×10^{-10}	[7-4]
Loss tangent, tan δ	5×10^{-4}	[7-4]

Retardation in Various Electro-Optic Materials. The general form of the relative phase retardation Γ is

$$\Gamma = \frac{2\pi\ell}{\lambda} (N_y - N_x)$$

where ℓ is the length of the crystal traversed by the light beam, λ is the free space wavelength and N_y and N_x are the principal refractive indices of the induced birefringement element. The explicit relations for the various electro-optic materials are shown in Table 7-2.

7.2.2 Design of Intensity Modulators

The customary and only practical mode of optical modulation using KDP and its isomorphs is that which applies the modulating signal field along the crystalline Z-axis (optic axis). If it is desired to avoid the natural birefringence and at the same time realize maximum electro-optic phase retardation, the light beam must also travel parallel to the Z-axis. Under the influence of a positive signal field, the principal axes of induced birefringence in the X-Y plane lies at ±45 degrees relative to the crystalline axes, with the X-axis being the fast axis, i.e., $N_x < N_y$. In KDP-type crystals

Crystal Class or Material	Retardation Equation	Examples	Indices of Refraction	Electro-Optic Coefficients
$\bar{4}2$ m	$\Gamma = \frac{2\pi\ell}{\lambda} N_o^3 r_{63} E$	KDP	$N_o = 1.51$	$r_{63} = 1.03 \times 10^{-11}$ m/volt
		KD*P	$N_o = 1.51$	$r_{63} = 2.6 \times 10^{-11}$ m/volt
$\bar{4}3$ m	$\Gamma = \frac{2\pi\ell}{\lambda} N_o^3 r_{41} E$	Cu Cl	$N_o = 1.93$	$r_{41} = 6.14 \times 10^{-12}$ m/volt
3 m	$\Gamma = \frac{2\pi\ell}{\lambda} N_o^3 r_{22} E$ (field parallel to b-axis)	$LiNbO_3$	$N_o = 2.237$ $N_e = 2.157$ (at 1 micron)	$r_{22} = 6.7 \times 10^{-12}$ m/volt $\left(\frac{N_e}{N_o}\right)^3 r_{33} - r_{13} = 1.73 \times 10^{-11}$ m/volt
	$\Gamma = \frac{\pi\ell}{\lambda} N_o^3 \left[\left(\frac{N_e}{N_o}\right)^3 r_{33} - r_{13}\right] E$ (field parallel to c-axis)	$LiTaO_3$	$N_o = 2.139$ $N_e = 2.143$	r_{22} not yet available $\left(\frac{N_e}{N_o}\right)^3 r_{33} - r_{13} = 1.73 \times 10^{-11}$ m/volt
m3m	$\Gamma = \frac{\pi\ell}{\lambda} N_o^3 (g_{11} - g_{12}) \epsilon^2 E^2$	KTN	$N_o = 2.29$	$(g_{11} - g_{12})\epsilon^2 = 1.36 \times 10^{-15}$ (m/volt)^2
Kerr Fluids	$\Gamma = 2\pi B \ell E^2$	Nitrobenzene	$N_o = 1.55$	$B = 2.5 \times 10^{-12}$ (m/volt)^2

N_o is the ordinary refractive index

N_e is the extraordinary refractive index

The electro-optic coefficients for the crystals are the "unclamped", zero stress values obtained with dc or low frequency fields.

g_{11} and g_{12} are the quadratic electro-optic coefficients in the cubic axes system.

ϵ is the permitivity

B is the Kerr constant

Table 7-2. Properties of some typical electro-optic materials.

$$N_x = N_o - \frac{1}{2} N_o^3 r_{63} E_m, \quad N_y = N_o + \frac{1}{2} N_o^3 r_{63} E_m \tag{7-4}$$

so that (7-2) is applicable when $r = r_{63}$. The field strength E_m in terms of the voltage V_m is, however, $E_m = V_m/\ell$, and (7-2) becomes

$$\Gamma = \frac{2\pi N_o^3 r_{63} V_m}{\lambda} . \tag{7-5}$$

The output phase difference, Γ, of optical components polarized along the X and Y axes, is therefore independent of the length of the crystal. To reach the maximum $\Gamma(= \pi)$ in KDP, for light of wavelength $\lambda = 6328$ Å, requires a rf signal voltage $V_m = 11$ kv, peak to peak, which is obviously much too high for practical operation.

We shall discuss two types of modulators which are designed to circumvent the above limitation. Both systems provide a long interaction length with modulating fields sufficiently strong to yield ample modulation depth at low levels of driving power.

Standing Wave or Video Intensity Modulators

The first modulator is a multielement device consisting of a linear chain of small Z-cut KDP crystals separated by thin, flat-ring electrodes. This is shown schematically in Figure 7-1. Similar modulators have already been constructed and operated successfully in the Hughes Research Laboratories. Figure 7-2 is a photograph of one such unit containing 16 elements. Alternate electrodes carry common signal voltage polarity; therefore, adjacent crystal elements have opposing fields. Since according to (7-4) the fast and slow axes of induced birefringence exchange roles upon reversal of the field, alternate crystal elements are rotated 90 degrees about their common Z-axis, so that the instantaneous fast and slow axes all have a common direction throughout the chain. The effect of this arrangement is to produce a cumulative phase shift between x and y components of the light beam which is the sum of the separate elemental phase shifts.

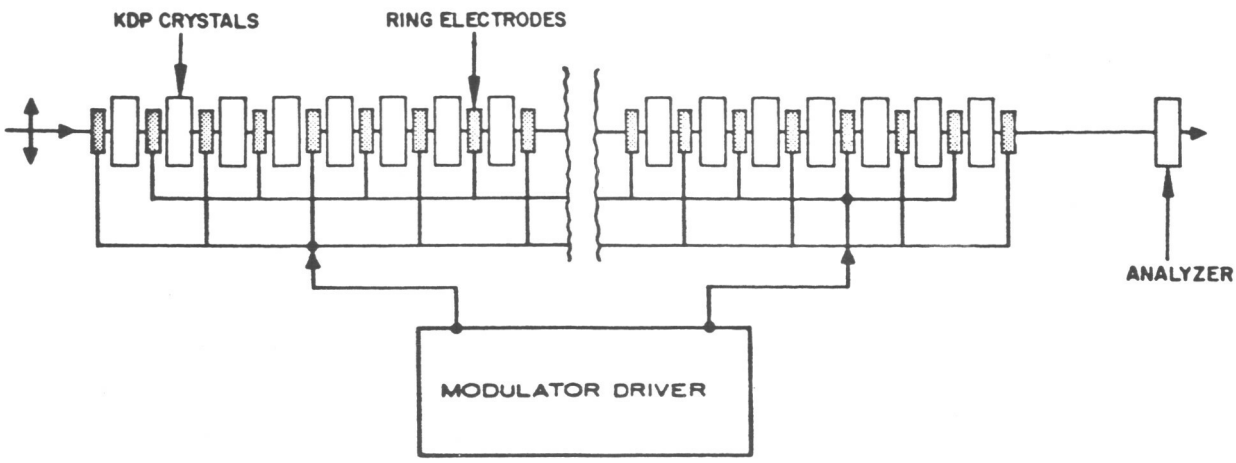


Figure 7-1. Multielement optical modulator.

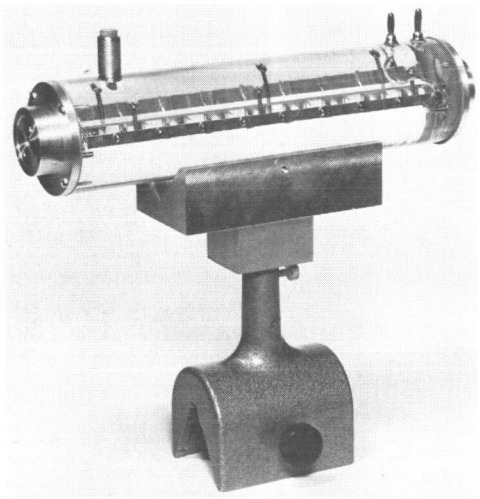


Figure 7-2. Sixteen-element KDP crystal modulator. (HAC photo M4274)

By the same token, the driving voltage required to produce a given total phase shift is 1/n times the voltage needed to produce the same phase shift in a single element, where n is the number of elements.

The second, or alternative, modulator design has quite a different configuration.[5] It consists of two long, narrow, parallel-strip condensers filled with KDP. These are made as nearly identical as possible and are oriented in tandem along the optical beam axis. The electric fields are applied as before along the crystalline optic axis; however, this direction is now transverse to the beam, and natural

birefringence must be taken into account. The purpose of having two similar units is to effect a cancellation of this natural birefringence by rotating the second unit 90 degrees relative to the first. Maximum optical phase shift is achieved if the X-axis of induced birefringence in the first unit is directed transverse to the beam and parallel to the plane of the electrodes, and if the Y-axis in the second unit is similarly oriented. The incident light beam is plane polarized at 45 degrees to the first X-axis. Figure 7-3 illustrates schematically this modulator system. A simplification in the design of this configuration can be made which allows the use of a single strip line.[2] The two crystal elements are aligned with their Z-axes parallel and are separated by a $\lambda/2$ plate, which is the optical equivalent of the original 90 degree rotation.

The optical phase shift Γ produced by this type of modulator is

$$\Gamma = \frac{2\pi}{\lambda} \left\{ \ell N_o^3 r_{63} E_m + \Delta\ell (N_e - N_o) \right\} \tag{7-6}$$

in which ℓ is the average value of the separate crystal lengths and $\Delta\ell$ is their length difference. A derivation of (7-6) is given below. An optical compensator,[2] as shown in Figure 7-3, can be incorporated into the modulator and adjusted to cancel out the residual birefringence resulting from $\Delta\ell$. Neglecting temperature effects for the moment, we may write the signal-dependent expression for Γ as

$$\Gamma = \frac{2\pi\ell N_o^3 r_{63} V_m}{\lambda t} \tag{7-7}$$

where again V_m is the signal voltage amplitude and t is the electrode, or plate, separation. Equation (7-7) is the same as (7-5) except for the factor ℓ/t. Such an enhancement of the field strength is a cogent reason for considering this form of modulator.

We shall analyze the resulting optical phase shift developed with the configuration illustrated in Figure 7-3, taking account of a slight difference in the lengths of crystal in the two elements. The incident laser beam is polarized at 45 degrees to the X- or fast-axis of the first element and to the Y-axis of the second element.

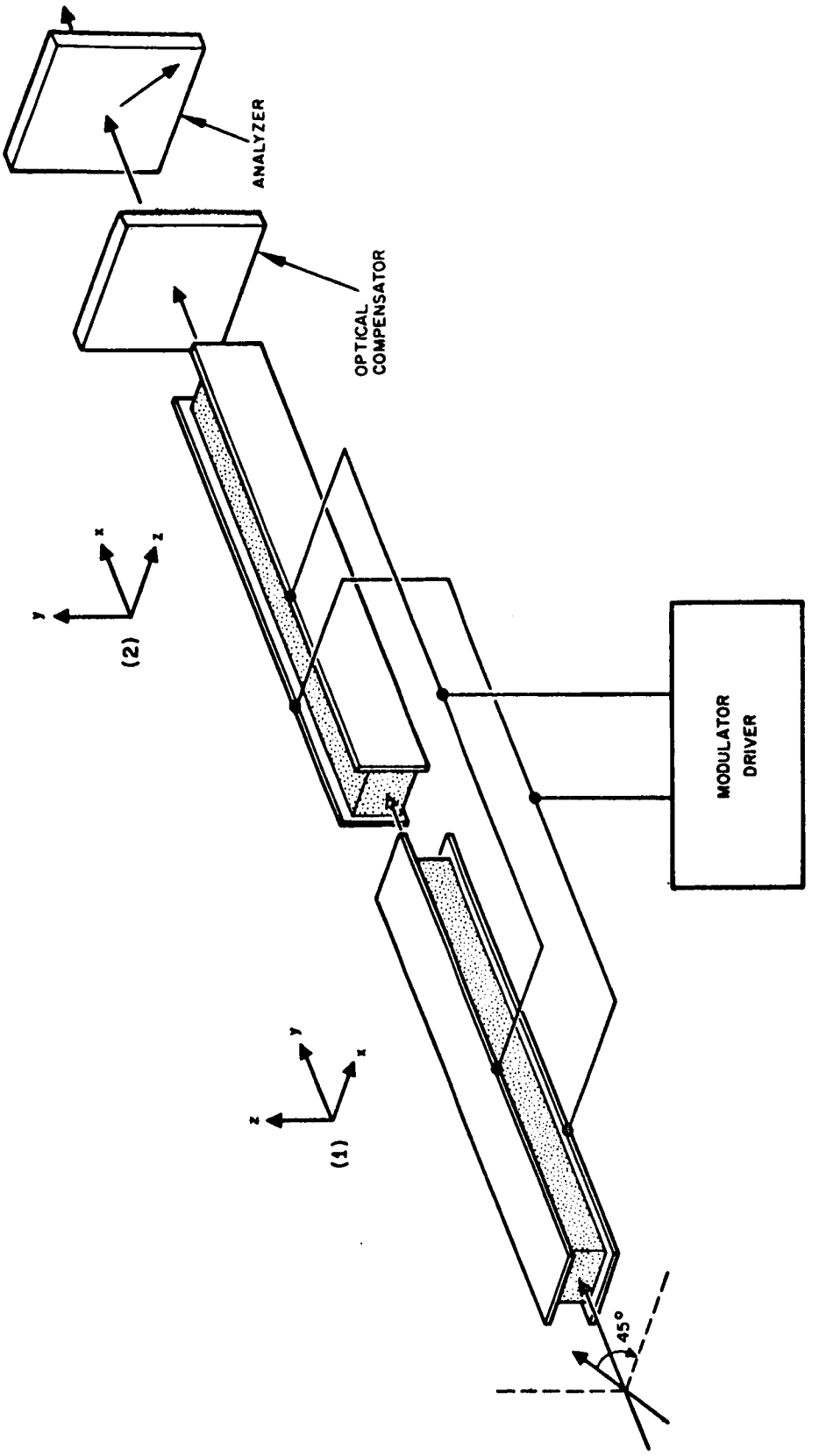


Figure 7-3. Two-element parallel-strip optical modulator.

Let

$\phi_v \equiv$ total phase retardation of vertical beam component

$\phi_h \equiv$ total phase retardation of horizontal beam component

$\ell_1 \equiv$ length of first modulator crystal

$\ell_2 \equiv$ length of second modulator crystal

$\ell = 1/2(\ell_1 + \ell_2)$

$\Delta\ell = |\ell_1 - \ell_2|$.

Then

$$\phi_v = \frac{2\pi}{\lambda}(\ell_1 N_z + \ell_2 N_y)$$

$$\phi_h = \frac{2\pi}{\lambda}(\ell_1 N_x + \ell_2 N_z),$$

and

$$\Gamma = \phi_v - \phi_h = \frac{2\pi}{\lambda}\left\{\ell_2 N_y - \ell_1 N_x + (\ell_1 - \ell_2)N_z\right\}.$$

Using (7-4) and $N_z = N_e$, we have

$$\Gamma = \frac{2\pi}{\lambda}\left\{\ell N_o{}^3 r_{63} E_m + \Delta\ell(N_e - N_o)\right\}. \qquad (7\text{-}6)$$

The temperature dependence of the static birefringence is contained in the second member of (7-6). Applying (7-3) with ℓ replaced by $\Delta\ell$, we obtain

$$\frac{\Delta\Gamma}{\Delta\Theta} = \frac{2\pi\Delta\ell}{\lambda}k.$$

With the KDP value of k from Table 7-1, this gives

$$\frac{\Delta\Gamma}{\Delta\Theta} = 109\,\Delta\ell \text{ rad/}^\circ\text{C } (\Delta\ell \text{ in meters}).$$

No technological difficulty is anticipated in achieving a $\Delta\ell$ considerably less than 0.10 mm. Thus we can assume

$$\frac{\Delta\Gamma}{\Delta\Theta} < 0.01 \text{ rad/}^\circ\text{C}.$$

Traveling Wave Intensity Modulator

An optical intensity modulator is related to the phase modulator which is directly adaptable to intensity modulation by suitable choice of input polarization and inclusion of a linear polarizing analyzer at the output. Moreover all the advantages possessed by traveling wave structures for broadband phase modulation at microwave frequencies apply with equal force to broadband intensity modulation.

Figure 7-4 illustrates the basic configuration of a strip line traveling-wave optical intensity modulator using a cubic crystal of

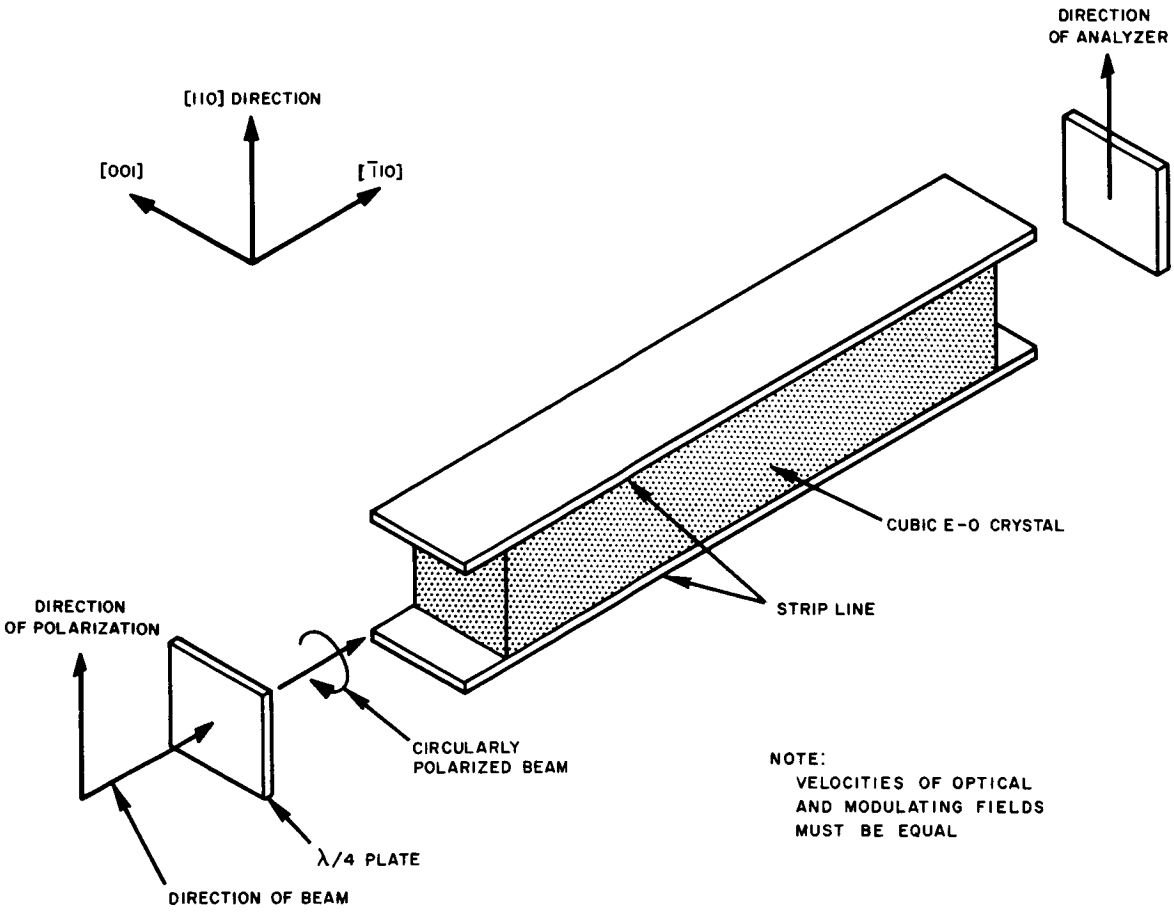


Figure 7-4. Strip-line traveling-wave modulator configuration.

$\bar{4}3m$, such as cuprous chloride. We have chosen this crystal type for illustration as representing the ideal, since natural birefrigence is absent, although it must be remembered that CuCl and other similar electro-optic crystals are difficult to obtain.

The indicated orientation of the crystalline axes represents the optimum configuration in the intensity-modulation mode of operation. The incident vertically-polarized laser beam is transformed by a quarter-wave plate into circular polarization in order to set the proper optical bias for best linear operation. By the action of a vertical modulation field along the [110] direction, the original circular polarization is rendered elliptical. That part of the beam which emerges from the vertically-oriented analyzer exhibits the desired intensity modulation according to the impressed signal. It can be shown that the transmitted light intensity I, relative to the laser intensity I_o, is

$$I = 1/2\ I_o\ (1 + \sin\ \Gamma)$$

where Γ is the relative optical phase shift of the two principal beam components within the crystal. For a CW modulation signal $\Gamma = \Gamma_o \sin\ \omega_m t$ and if $\Gamma_o \leq .5$ rad., we may let $\sin\ \Gamma \cong \Gamma$, so that

$$I = 1/2\ I_o\ (1 + \Gamma_o \sin\ \omega_m t)$$

Thus Γ_o represents the usual modulation index m.

The KDP phase modulator can also be used as an amplitude modulator[2] provided a suitable optical compensator is added to the system to counteract the natural birefringence. This compensator could for example replace the quarter-wave plate. The initial incident beam and final analyzer should be oriented at 45 degrees to the vertical.

A disadvantage of this scheme is that the natural birefringence of ADP and KDP is temperature dependent. Thus small fluctuations in the crystal temperature would cause a variation in the optical bias and result in distortion of the output. This problem can, however, be

largely overcome by using two identical modulators in tandem and rotated 90 degrees with respect to each other. If the [110] direction in the second modulator crystal is oriented in the transverse plane and parallel to the strips, the natural birefringence is cancelled. Of course, care must be taken to maintain a common temperature in each unit.

Dielectric Power Absorption

The power absorption per unit volume of a dielectric upon which an electric field E is impressed is given by

$$P = \sigma \frac{E^2}{2}$$

where σ is the dielectric conductivity; σ is defined as the product of the signal angular frequence ω_m and the loss factor ϵ'' of the material. The loss tangent of the material is defined by the relation

$$\tan \delta = \frac{\epsilon''}{\epsilon'}$$

where ϵ' is the real part of the permittivity of the material. The total power, P, absorbed by a dielectric with cross-sectional area, A, and length, ℓ, is then

$$P = \frac{1}{2} \omega_m \epsilon' E^2 A \ell \tan \delta$$

7.2.3 Design of Optical Frequency Translators

In practice a laser heterodyne communication link normally employs two identical lasers, one being translated in frequency. Either could be used as the local oscillator, with the other serving as transmitter carrier. Optical frequency translation can be affected by several techniques.

One method of optical frequency translation for SSBSC operation is the electro-optic rotating birefringent plate.[6, 7] In some forms, this instrument is capable of completely transforming a circularly polarized light beam into a beam of the opposite polarity and shifted in frequency by twice the frequency of rotation of the wave plate. More generally, however, only a portion of the original beam is thus shifted, and suppression of the carrier is affected by a circular polarizer. Cubic crystals of symmetry, class $\bar{4}3m$ and m3m, when driven by a rotating electric field in their (111) crystal plane, exhibit a true rotating birefringence in that plane, and thus produce frequency translation of a beam of light traveling along the [111] direction. The most popular and readily available electro-optic crystals, of class $\bar{4}2m$ such as KDP, do not have the above property; and thus an alternate technique is used with them.[8, 9, 10, 11] Two crystals are oriented in tandem along the optic beam and driven in phase quadrature. These elements are oriented with their principal axes of induced birefringence at an angle of 45 degrees with respect to each other. For incident circularly-polarized light this arrangement simulates a rotating birefringent plate, producing a component of circular polarization of the opposite sense and shifted by the signal frequency.

The chief advantage of the rotating wave plate method is its versatility in allowing simultaneous frequency translation and amplitude, frequency, or FSK modulation of the shifted carrier. The two basic types reviewed briefly here will be discussed more fully in the following pages. A detailed mathematical analysis of rotating birefringent elements is given.

Single-Element Frequency Translators

The one-element SSBSC modulator constitutes a true rotating birefringent element which serves the function of optical frequency translation by adding a constant rate of change of phase to a wave of given input frequency. It is thus the optical analog of the conventional microwave phase shifter which utilizes a mechanically rotated birefringent half-wave plate.

To realize such an optical device requires in practice the use of electro-optic crystals of special symmetry classes: $\bar{4}3$ m, such as cuprous chloride, zinc sulfide, and hexamine (HMTA), 3m, such as $LiNbO_3$, or m3m, such as $KTa_{0.65}Nb_{0.35}O_3$ (KTN) or other similar perovskites operating in their paraelectric phase. The first two classes possess a first order electro-optic (Pockels) effect, the third, a quadratic (Kerr) effect. For classes $\bar{4}$3m and m3m the crystals are driven by an electric field rotating in the crystalline (111) plane, while the optical beam, circularly-polarized, travels through in the [111] direction. For class 3m, the field is in the (001) plane; the beam direction is [001]. Under the influence of a signal frequency f_m rotating, say, clockwise, the induced axes of birefringence of the m3m type rotate with the field at frequency f_m, while those of the $\bar{4}$3m and 3m type rotate counterclockwise at a frequency $1/2\ f_m$. Since, as noted previously, the optical beam undergoes a frequency translation equal to twice the rotation rate of the wave plate, this shift can equal either f_m or $2f_m$ depending upon the class of crystal employed. In either case, if the amplitude of the applied signal voltage is sufficient to produce a full half-wave of induced birefringence, the optical frequency translation is complete.

Unfortunately crystals of the types mentioned above are difficult to synthesize in large sizes and of high optical quality, so that construction of single-element SSBSC modulators is limited and difficult. For this reason, we discuss an alternative device in the following section which employs the more readily available material KDP or its isomorphs, in a two-element simulated rotating wave plate. We indicate, however, in Figures 7-5 and 7-6, respectively, the low-frequency and microwave schematics of a one-element device.

Double-Element Frequency Translators

The necessity of employing the two-element modulator for optical frequency translation is dictated mainly by the crystal characteristics of the more readily available and high quality materials of class $\bar{4}$2m, such as potassium dihydrogen phosphate (KDP) and its isomorphs.

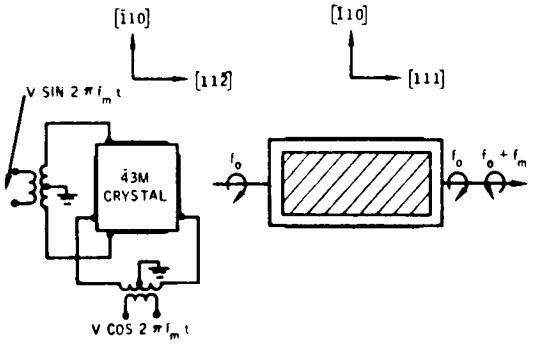


Figure 7-5. Low frequency one-element optical frequency translator.

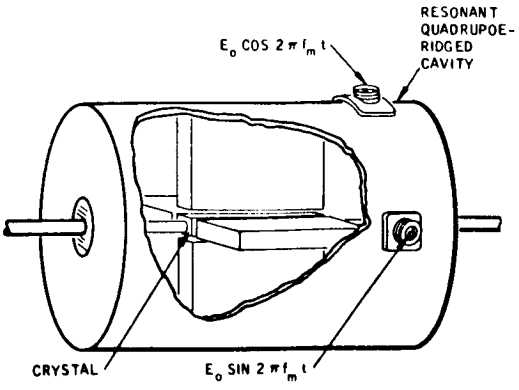


Figure 7-6. A possible microwave optical frequency trnaslator.

Because this type is naturally birefringent, it is customary to orient the light beam along the optic axis, in order to avoid this unwanted birefringence in the absence of modulating signal*. For optimum electro-optic effect it is also necessary to apply the signal field along the c-axis; thus the beam and signal field are parallel. With this configuration the induced axes of birefringence lie in the (001) plane at a fixed orientation of 45 degrees relative to the a and b axes, exchanging roles of "fast" and "slow" axis upon reversal of the signal field.

The action of a simulated rotating birefringent plate is provided by orienting two identical crystal elements in tandem along the beam axis and so oriented relative to each other that the corresponding a and b axes (or equivalently the induced principal axes) lie at 45 degrees with respect to each other. The modulating signal is applied to the two elements in phase quadrature. This configuration behaves like a single wave plate rotating at half the frequency of the applied signal, so that the optical frequency is translated by an amount equal to that of the signal. Figure 7-7 illustrates this particular system under low

*A traveling wave modulator, described elsewhere, operates with light beam perpendicular to the optic axis but has built-in cancellation of natural birefringence.

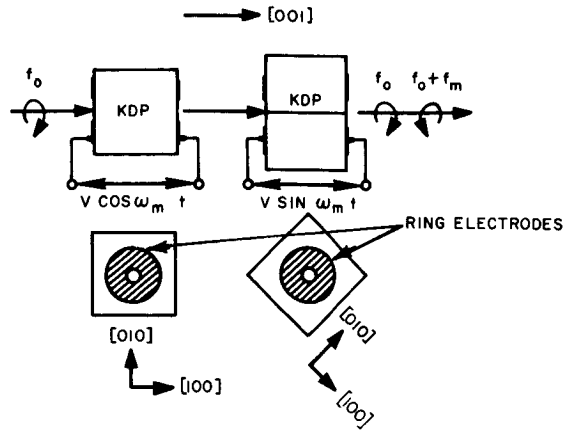


Figure 7-7. Two element KDP optical frequency translator for SSBSC modulator.

frequency operation. At microwave frequencies, the simple capacitative electrodes and associated circuit is normally replaced by suitable reentrant or cylindrical cavities. In some cases the crystals are driven in phase but separated by a distance equal to a quarter of a signal wavelength so that the optical photons automatically experience the quadrature driving fields by virtue of their transmit time.

It is shown in a later paragraph entitled Analysis of Frequency Translators that complete disappearance of the carrier occurs not at half-wave voltage when the optical relative phase retardation Γ_o is $\pi/2$ but rather when $J_o(\Gamma_o/\sqrt{2}) = o$ or when $\Gamma_o = 3.40$. Here J_o is the Bessel Function of zero order. Also, in addition to the first order single sideband, an upper and lower second-order sideband are generated; also a single third-order single sideband, etc. The optical amplitude of these sidebands are proportional to their respective Bessel Functions of the argument $\Gamma_o/\sqrt{2}$.

Analysis of Frequency Translators

The effect of various optical elements upon a collimated monochromatic light beam of known initial polarization can be conveniently analyzed by means of an operator calculus in which the electric vector of the light is represented by a two-element column matrix

$$\mathcal{E} = \begin{pmatrix} X \\ Y \end{pmatrix} e^{i\omega t}$$

and the optical element is represented by a 2 by 2 matrix operator

$$M = \begin{pmatrix} m_1 & m_4 \\ m_3 & m_2 \end{pmatrix}$$

The components X and Y are in general complex and completely define the state of polarization of the light in terms of some chosen space system of axes x, y normal to the direction of light propagation. The four components of M are likewise in general complex. Upon passing through a given optical element, the light vector undergoes a linear transformation symbolized by

$$\mathcal{E} = M\mathcal{E}_0$$

The optical element may be a birefringent plate, an anisotropic absorber, or a rotator. If more than one optical element is present in the system, the resultant M is simply a matrix product,

$$M = M_n M_{n-1} \cdots\cdots M_2 M_1$$

of the individual element matrices M_i. The elements are numbered in the order in which the beam traverses them. This section is concerned only with birefringent plates, static and induced, and with anisotropic absorbers (polarizers or analyzers).

Explicit expressions for the M matrices will depend upon the coordinate axes chosen as well as upon the physical properties of the associated optical element. A birefringent plate, for example, whose principal axes of retardation coincide with the space-fixed x, y axes takes the form

$$M = \begin{pmatrix} e^{i\frac{\phi}{2}} & 0 \\ 0 & e^{-i\frac{\phi}{2}} \end{pmatrix}$$

where ϕ is the relative phase retardation angle and the x-axis is assumed to be the fast axis. If the plate is so oriented that its principal x-axis makes an angle θ with the space x-axis, then in terms of the space-fixed system

$$M = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} e^{i\frac{\phi}{2}} & 0 \\ 0 & e^{-i\frac{\phi}{2}} \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

$$= \begin{pmatrix} e^{i\frac{\phi}{2}}\cos^2\theta + e^{-i\frac{\phi}{2}}\sin^2\theta & e^{i\frac{\phi}{2}}\cos\theta\sin\theta - e^{-i\frac{\phi}{2}}\cos\theta\sin\theta \\ e^{i\frac{\phi}{2}}\cos\theta\sin\theta - e^{-i\frac{\phi}{2}}\cos\theta\sin\theta & e^{i\frac{\phi}{2}}\sin^2\theta + e^{-i\frac{\phi}{2}}\cos^2\theta \end{pmatrix}$$

In the case of a uniformly rotating birefringent plate, θ is a linear function of time.

Applying the operator M of the general birefringent plate given above to the electric field vector of, say, left-circularly-polarized incident light, we obtain for the emergent light vector

$$\mathcal{E} = M\mathcal{E}_0 = (M) \begin{pmatrix} \frac{i}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} e^{i\omega t}$$

whence

$$X = \frac{1}{\sqrt{2}} \left\{ i\cos\frac{\Gamma}{2} - \sin\frac{\Gamma}{2} e^{-2i\theta} \right\}$$

$$Y = \frac{1}{\sqrt{2}} \left\{ \cos\frac{\Gamma}{2} - i\sin\frac{\Gamma}{2} e^{-2i\theta} \right\}$$

or

$$\mathcal{E}_x = \frac{1}{\sqrt{2}} \left\{ \cos \frac{\Gamma}{2} e^{i(\omega t + \frac{\pi}{2})} - \sin \frac{\Gamma}{2} e^{i(\omega t - 2\theta)} \right\}$$

$$\mathcal{E}_y = \frac{1}{\sqrt{2}} \left\{ \cos \frac{\Gamma}{2} e^{i\omega t} - \sin \frac{\Gamma}{2} e^{i(\omega t - 2\theta + \frac{\pi}{2})} \right\}$$

where Γ is the induced relative phase retardation or birefringence. The emergent beam thus consists of two parts: the original left-circularly-polarized beam of amplitude $\cos \Gamma/2$ with frequency $\omega/2\pi$, and a right-circularly-polarized component of amplitude $\sin \Gamma/2$ with frequency $1/2\pi(\omega - 2d\theta/dt)$.

Cubic crystals of symmetry class $\bar{4}3m$, such as CuC_1, ZnS, and HMTA, exhibit a first order electro-optic effect. Under the influence of a signal field rotating at frequency $\omega_m/2\pi$ in the (111) crystallographic plane, these materials act as true rotating birefringent plates. If the signal voltage components are given by

$$V_x = V_0 \cos \omega_m t \qquad V_y = V_0 \sin \omega_m t,$$

corresponding to counter-clockwise rotation of the impressed field, then it can be shown that the principal axes of the optical indicatrix rotate <u>clockwise</u> in this plane at an angular velocity $d\theta/dt = -1/2\ \omega_m$. Thus an incident left-circularly-polarized beam entering the crystal along the [111] axis will, upon emerging, contain a right-circularly-polarized upper sideband at frequency $(1/2\pi)(\omega + \omega_m)$.

Cubic crystals of class m3m, such as KTN, show only a second order electro-optic effect. It turns out that with an applied field in the (111) plane, the fast axis (x-axis) of the optical indicatrix in this plane is always parallel to the applied field regardless of its direction. Therefore, a rotating signal field will give rise to a birefringent plate rotating in the same direction at the same frequency. If again we choose

$$V_x = V_0 \cos \omega_m t \qquad V_y = V_0 \sin \omega_m t,$$

then a left-circularly-polarized beam incident along the [111] direction will emerge containing a right-circular lower sideband at frequency $(1/2\pi)(\omega - 2\omega_m)$. Moreover it is apparent that the signal frequency has been doubled. It is also apparent that in both crystal systems just described, an upper or lower sideband can be chosen at will by properly phasing the components of the signal voltage.

Another important type of electro-optic crystal is that belonging to the symmetry class $\bar{4}2$ m, most familiar among which are KDP and ADP. These crystals are naturally birefringent, possess a first-order electro-optic effect, and for maximum sensitivity must be operated with applied field parallel to the traversing light beam and parallel to the crystallographic c-axis (optic axis). In this case the induced principal axes of the optical indicatrix in the x-y plane lie at a constant angle of 45 degrees relative to the crystallographic a and b axes. For this reason, a true rotating birefringent plate cannot be realized, and a different approach is used which provides an "effective" rotating plate. The technique employs two identical crystals in tandem, one being rotated about the c-axis 45 degrees relative to the other. Signal voltages of equal amplitude are impressed along the c-axes in phase quadrature.

Let crystal No. 1 be oriented with its crystallographic a-axis parallel to the space x-axis, and let crystal No. 2 be rotated clockwise 45 degrees so that its induced principal x-axis is along x. Left-circularly-polarized light is incident on the first crystal along its c-axis. The matrix operator which describes the properties of the emergent beam is then the product M_2M_1 of the separate matrices associated with the two crystals. In this case we have

$$M_1 = \begin{pmatrix} \cos\frac{\Gamma_1}{2} & i\sin\frac{\Gamma_1}{2} \\ i\sin\frac{\Gamma_1}{2} & \cos\frac{\Gamma_1}{2} \end{pmatrix}, \quad M_2 = \begin{pmatrix} e^{i\frac{\Gamma_2}{2}} & 0 \\ 0 & e^{-i\frac{\Gamma_2}{2}} \end{pmatrix}$$

where Γ_1, Γ_2 are the relative phase retardations of the crystals. We may write

$$\Gamma_1 = \Gamma \cos \omega_m t , \qquad \Gamma_2 = \Gamma \sin \omega_m t$$

and note that the time dependence of the matrices resides now in Γ rather than θ. This system is, however, essentially equivalent to a rotating birefringent plate of angular frequency $1/2\,\omega_m$. The electric vector of the emergent beam is given by

$$\mathcal{E} = M_2 M_1 \mathcal{E}_0 = \begin{pmatrix} e^{i\frac{\Gamma_2}{2}} \cos\frac{\Gamma_1}{2} & ie^{i\frac{\Gamma_2}{2}} \sin\frac{\Gamma_1}{2} \\ ie^{-i\frac{\Gamma_2}{2}} \sin\frac{\Gamma_1}{2} & e^{-i\frac{\Gamma_2}{2}} \cos\frac{\Gamma_1}{2} \end{pmatrix} \begin{pmatrix} \frac{i}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} e^{i\omega t}$$

This yields

$$X = \frac{1}{2}\left\{ e^{i\left[\frac{\Gamma_0}{\sqrt{2}} \sin\left(\omega_m t + \frac{\pi}{4}\right) + \frac{\pi}{4}\right]} - e^{i\left[\frac{\Gamma_0}{\sqrt{2}} \sin\left(\omega_m t - \frac{\pi}{4}\right) - \frac{\pi}{4}\right]} \right\}$$

$$= \frac{1}{2} e^{i\frac{\pi}{4}} \sum_{n=-\infty}^{\infty} J_n\left(\frac{\Gamma_0}{\sqrt{2}}\right) e^{in\left(\omega_m t + \frac{\pi}{4}\right)} - \frac{1}{2} e^{-i\frac{\pi}{4}} \sum_{n=-\infty}^{\infty} J_n\left(\frac{\Gamma_0}{\sqrt{2}}\right) e^{in\left(\omega_m t - \frac{\pi}{4}\right)}$$

$$Y = \frac{1}{2}\left\{ e^{-i\left[\frac{\Gamma_0}{\sqrt{2}} \sin\left(\omega_m t + \frac{\pi}{4}\right) + \frac{\pi}{4}\right]} + e^{-i\left[\frac{\Gamma_0}{\sqrt{2}} \sin\left(\omega_m t - \frac{\pi}{4}\right) - \frac{\pi}{4}\right]} \right\}$$

$$= \frac{1}{2} e^{-i\frac{\pi}{4}} \sum_{n=-\infty}^{\infty} J_n\left(-\frac{\Gamma_0}{\sqrt{2}}\right) e^{in\left(\omega_m t + \frac{\pi}{4}\right)} + \frac{1}{2} e^{i\frac{\pi}{4}} \sum_{n=-\infty}^{\infty} J_n\left(-\frac{\Gamma_0}{\sqrt{2}}\right) e^{in\left(\omega_m t - \frac{\pi}{4}\right)}$$

Expanding the Bessel series and recalling the properties that

$J_n(x) = (-1)^n J_n(-x)$ and $J_n(x) = (-1)^n J_{-n}(x)$,

we get finally for the components of the emergent light vector:

$$\mathcal{E}_x = \frac{1}{\sqrt{2}} J_0\left(\frac{\Gamma_0}{\sqrt{2}}\right) e^{i\left(\omega t + \frac{\pi}{2}\right)} + J_1\left(\frac{\Gamma_0}{\sqrt{2}}\right) e^{i\left[(\omega + \omega_m)t + \frac{\pi}{2}\right]}$$

$$+ \frac{1}{\sqrt{2}} J_2\left(\frac{\Gamma_0}{\sqrt{2}}\right)\left[e^{i\left[(\omega + 2\omega_m)t + \frac{\pi}{2}\right]} - e^{i\left[(\omega - 2\omega_m)t + \frac{\pi}{2}\right]}\right] + \cdots\cdots,$$

$$\mathcal{E}_y = \frac{1}{\sqrt{2}} J_0\left(\frac{\Gamma_0}{\sqrt{2}}\right) e^{i\omega t} - J_1\left(\frac{\Gamma_0}{\sqrt{2}}\right) e^{i(\omega + \omega_m)t}$$

$$+ \frac{1}{\sqrt{2}} J_2\left(\frac{\Gamma_0}{\sqrt{2}}\right)\left[e^{i(\omega + 2\omega_m)t} - e^{i(\omega - 2\omega_m)t}\right] + \cdots\cdots,$$

We thus have generated, in addition to the original left-circularly-polarized beam, a right-circularly-polarized upper sideband at frequency $(1/2\pi)(\omega + \omega_m)$, a left-circularly-polarized upper and lower sidebands of equal amplitude at frequency $(1/2\pi)(\omega \pm 2\omega_m)$, etc. The normalized amplitudes of the first three sidebands as a function of applied voltage are shown in Figure 7-8.

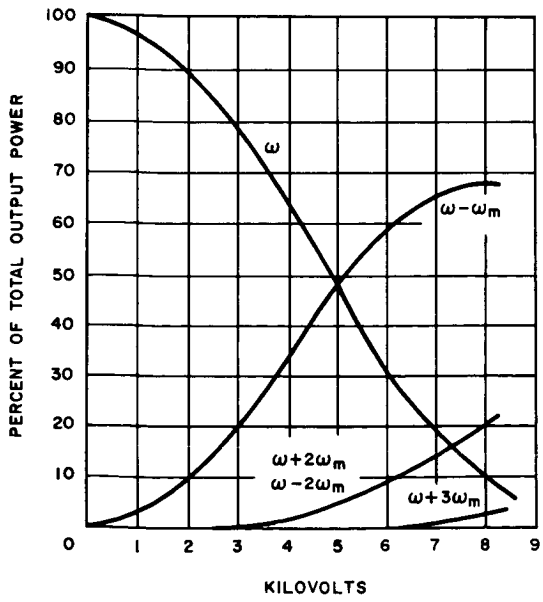


Figure 7-8.

Frequency components at the output of a two-element frequency translator.

Similar results are obtainable with a single KDP modulator if the second crystal is replaced by $\lambda/8$ plate and mirror. The distance from modulator to mirror is so adjusted that the round trip photon transit time equals one-quarter period of the modulating signal. The KDP is again oriented with its crystalline axes parallel to the space x-y axes, and the eighth-wave plate has its principal axes of birefringence likewise parallel to the space axes – that is, at 45 degrees to the induced principal axes of the modulator. The $\lambda/8$ plate serves to simulate the 45 degree rotation of the second crystal. In this arrangement, contrary to those described previously, the incident beam should be vertically plane polarized. After its second passage through the crystal the reflected beam contains a single upper sideband horizontally polarized.

This arrangement is equivalent to two identical modulators operating in phase quadrature and separated by a quarter-wave plate. The appropriate matrices are therefore

$$M_1 = \begin{pmatrix} \cos\frac{\Gamma_1}{2} & i\sin\frac{\Gamma_1}{2} \\ i\sin\frac{\Gamma_1}{2} & \cos\frac{\Gamma_1}{2} \end{pmatrix}$$

$$M_2 = \begin{pmatrix} e^{i\frac{\pi}{4}} & 0 \\ 0 & e^{-i\frac{\pi}{4}} \end{pmatrix}$$

$$M_3 = \begin{pmatrix} \cos\frac{\Gamma_2}{2} & i\sin\frac{\Gamma_2}{2} \\ i\sin\frac{\Gamma_2}{2} & \cos\frac{\Gamma_2}{2} \end{pmatrix}$$

The transmitted beam is obtained by performing the operation

$$\mathcal{E} = M_3 M_2 M_1 \mathcal{E}_0$$

Neglecting Bessel terms of higher order than 1, the results of this operation give, for the components of the transmitted beam,

$$\mathcal{E}_x = \sqrt{2}\, J_1\left(\frac{\Gamma}{\sqrt{2}}\right) e^{i\left[(\omega + \omega_m)t + \frac{\pi}{4}\right]}$$

$$\mathcal{E}_y = J_0\left(\frac{\Gamma}{\sqrt{2}}\right) e^{i(\omega t - \frac{\pi}{4})}$$

The emergent beam thus contains no x-component at the carrier frequency and no y-component at the upper sideband.

7.2.4 Design of Phase Modulators

Traveling wave modulators characteristically possess wide-band capabilities and may take any of several forms. Perhaps the most promising and most efficient TW modulator is the parallel strip transmission line version,[12] first constructed using ADP (ammonium dihydrogen phosphate) and operated as a pure optical phase modulator. A cross-sectional view of this type is shown schematically in Figure Figure 7-9. The strip line is filled along its center with ADP crystals

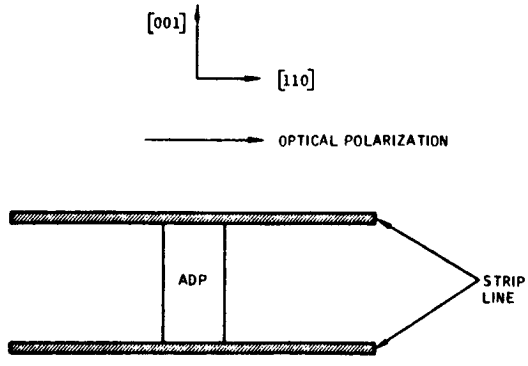


Figure 7-9. Cross sectional view of parallel strip transmission line ADP phase modulator.

oriented as shown, with one of the principal axes, [110], of induced birefringence in the transverse plane and parallel to the plane of the strips. For operation as a pure optical phase modulator, the optical polarization is also oriented in this direction, and the natural birefringence of the crystal is only of consequence if the crystalline axes are not perfectly oriented relative to the laser beam.

We note that in this configuration the applied signal field is now transverse to the beam, and thus the optical phase retardation is directly proportional to the length of the transmission line. A chief advantage of this structure, in addition to the naturally broadband feature of strip lines, is that it can be made quite long. Thus efficient operation is possible with very modest power requirements.

In high frequency applications of strip line modulators, a necessary criterion for optimum performance is that the phase velocity of the microwave signal traveling down the transmission line be matched to that of the light beam within the crystal dielectric. Since the dielectric constant of most crystalline materials is appreciably higher for microwaves than for light, the velocity matching is accomplished by a proper choice of the dielectric filling factor, that is, the ratio of the dielectric width to that of the strip line.[12, 13]

The effective microwave index of refraction of the composite line is virtually indpendent of frequency up to a point where the principal mode of transmission is no longer TEM-like. At higher frequencies the microwave velocity becomes highly dispersive until a point is eventually reached where the power resides wholly within the dielectric. The mode eventually becomes once again TEM-like, and the velocity is that characteristic of the dielectric alone.[13] These facts place an upper limit on the usable frequency range which is determined by the ratio of the crystal and air dielectric. For practical dimensions, sufficiently large to accommodate laser beams of about 2 mm diameter, this frequency limit is about 3 Gcps for ADP.[12]

7.2.5 Design of Polarization Modulators

A discussion of this important form of optical modulation will be treated in subsequent reports. In particular, the design of an optical PCM/PL system will be given. As an example, the high data rate system developed by Hughes for NASA-MSC shown in Figure 7-10, will be described.

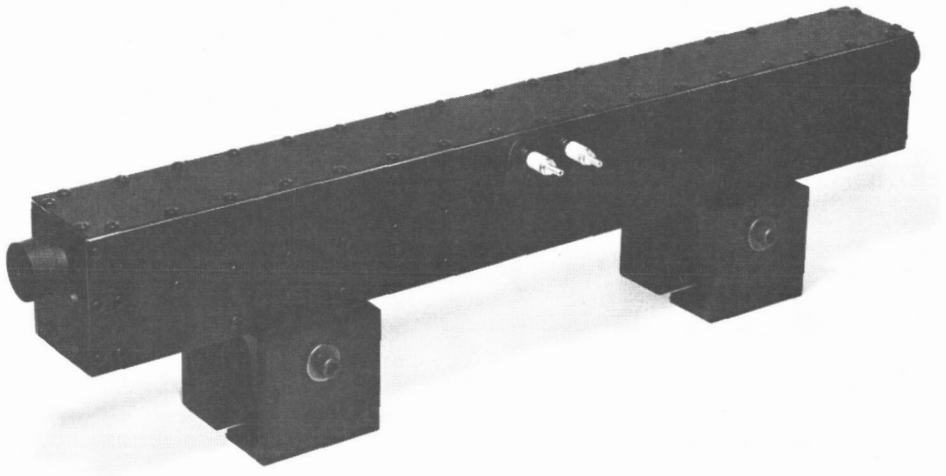


Figure 7-10. Modulator for PCM/PL system.
(HAC photo M4343)

7.3 ELASTO-OPTIC MODULATION

7.3.1 Theory of the Elasto-Optic Effect

The phenomenon of diffraction of light by the passage of a sound wave through the optical medium is largely governed by the change in refractive index induced in the medium by the passage of the sound wave and also by the radiation geometry of the sound and light wave sources. In this section we will list the equations which relate the change in refractive index to the stress-strain amplitude generated by the ultrasonic wave.

The elasto-optical coefficients for an anisotropic medium $p_{ijk\ell}$ are defined by[14, 15]

$$\Delta B_{ij} = p_{ijk\ell}\eta_{k\ell}$$

where η is the strain and B_{ij} is the dielectric impermeability tensor, which is related to the dielectric constant κ_{ij} by

$$\kappa_{ij}B_{jk} = \delta_{ik}$$

where δ_{ik} is the Kronecker delta. For the case of cubic crystals κ is isotropic and thus

$$\Delta\kappa_{ij} = -\kappa^2\Delta B_{ij} = -\kappa^2 p_{ijk\ell}\eta_{k\ell} \cdots$$

Furthermore, if one restricts the problem to the propagation of longitudinal sound waves (no shear components) along principal crystallographic directions only, the strain component can be represented by a single subscript and the change in dielectric constant can be simply represented by the following equation

$$\Delta\left(\frac{1}{\kappa}\right)_i = \sum_j p_{ij}s_j$$

where s_j, the strain tensor, is a special case of $\eta_{k\ell}$, and p is the fourth rank elasto-optic tensor.

Thus if one passes a longitudinal sound wave along the [1, 0, 0] crystallographic axis of a simple cubic crystal such as Silicon or Germanium and we label this the j = 1 direction, the strain subscript will be (s_1). For the case of light wave polarization parallel to the direction of sound propagation the elasto-optic tensor p_{11} will apply; for orthogonal polarization the tensor p_{21} will apply. So that for parallel and orthogonal polarization relative to the direction of propagation of a pure logitudinal [1, 0, 0] ultrasonic wave one obtains the following expressions for the change in refractive index.

$$\Delta\left(\frac{1}{\kappa}\right)_1 = p_{11} s_1 \qquad \text{(Parallel)}$$

$$\Delta\left(\frac{1}{\kappa}\right)_2 = p_{21} s_1 \qquad \text{(Perpendicular)} \tag{7-8}$$

Due to the symmetry of a cubic crystal the directions [0, 1, 0] and [0, 0, 1] have the same elasto-optic behavior. However for the case of sound propagation along the [1, 1, 1] crystallographic direction the elasto-optic constant is a linear combination of p_{11}, p_{21} and p_{44}, the later term being generated by shear stresses. For the case of a rotation of coordinates where the Z direction is the [1, 1, 1] crystallographic direction represented by j = 3', the Y direction lies in the [1, 1, 0] plane and is represented by j = 1', one finds for sound along j = 3' and light polarization along j = 1' and along j = 3',

$$p_{1'3'} = \frac{1}{3}p_{11} + \frac{2}{3}p_{12} - \frac{1}{3}p_{44} \qquad \text{perpendicular polarization}$$

$$p_{3'3'} = \frac{1}{3}p_{11} + \frac{2}{3}p_{12} + \frac{2}{3}p_{44} \qquad \text{parallel polarization} \tag{7-9}$$

In more complicated crystals, such as uniaxial or biaxial, the unabridged notation applies and can be used in an analogous manner. The preceeding relations indicate that the elasto-optic behavior of a single cubic crystal is completely described by three constants p_{11}, p_{12}, and p_{44}. These constants can only be measured by inducing strain in the crystal and determining the resultant change in index of refraction. Some limited measurements have been made on the elasto-optic constants of selected crystals in the visible region of the spectrum but not at infrared wavelengths. Currently, work is in progress at Hughes Research Laboratories and elsewhere to evaluate the elasto-optic behavior of various materials at 3.39 and 10.6 microns.

Using the relationship between index of refraction and dielectric constant ($\kappa = N^2$) one finds the change in refractive index from Equations (7-8) and (7-9) to be

$$\Delta N_1 = -\frac{N^3}{2} p_{11} s_1 \qquad [1, 0, 0] \text{ sound, parallel light}$$

$$\Delta N_2 = -\frac{N^3}{2} p_{21} s_1 \qquad [1, 0, 0] \text{ sound, perpendicular light}$$

$$\Delta N_1^1 = -\frac{N^3}{2}\left(\frac{1}{3}p_{11} + \frac{2}{3}p_{12} - \frac{1}{3}p_{44}\right) s_3^1 \qquad [1, 1, 1] \text{ sound, perpendicular light}$$

$$\Delta N_3^1 = -\frac{N^3}{2}\left(\frac{1}{3}p_{11} + \frac{2}{3}p_{12} + \frac{2}{3}p_{44}\right) s_3^1 \qquad [1, 1, 1] \text{ sound, parallel light}$$

In the case of non-cubic crystals the tensor relationships become more complicated and terms such as p_{13}, p_{31}, p_{14}, etc., appear. A few values of p_{ij} for single transparent crystals follow. The data is old and probably not highly accurate.

Lithium Niobate

$p_{41} = p_{42} = p_{56} = 0.136$

$p_{11} = p_{22} = 0.032$

$p_{13} = p_{23} = 0.069$

$p_{31} - p_{32} = 0.153$

$p_{21} = p_{12} = 0.063$

$p_{33} = 0.061$

Fused Quartz

$p_{21} = p_{12} = 0.250$

KRS 5

$p_{11} = 0.25$ $p_{12} = 0.21$ $p_{44} = 0.157$

MgO

$p_{11} = -0.32$ $p_{12} = -0.08$ $p_{44} = ?$

LiF

$p_{11} = 0.02$ $p_{12} = 0.128$ $p_{44} = -0.064$

KB_r

$p_{11} = 0.22$ $p_{12} = 0.171$ $p_{44} = -0.026$

CdS

$p_{11} \approx 0.2$ at 6328 Å $p_{11} = 0.127$ at 3.39 μ

7.3.2 Theory of Ultrasonic Diffraction of Light

It has been shown both theoretically and experimentally[16, 17, 18] that if the angle of incidence of a light wave relative to the wavefront of a plane sound wave obeys the following equation then the intensity in the -1 diffraction order is stronger that that of the +1 order and that under optimum conditions the intensity of the +1 order may be reduced to zero.

$$\lambda_L = 2\lambda_s \sin \frac{\theta}{2}$$

where λ_L and λ_s are the optical and acoustic wavelengths, respectively, and θ is the angle between the incident and the diffracted beams. This phenomenon may be most easily understood if one considers the layered nature of the sound wave acting as a Bragg reflection grating (see Figure 7-11). At an appropriate angular condition the multiply-scattered beams are phased to produce a maximum in one direction and a minimum in all other directions. If the ultrasonic transducer is sufficiently long, one finds that there are more scattering centers, or in other terms the grating is large enough to yield a large scattering cross section, and the intensity of the -1 order beam is strengthened at the expense of the other beams. Under the Bragg condition the intensity of the -1 order beam relative to the zero beam is given by

$$\frac{I_{-1}}{I_o} = \sin^2\left(\frac{v}{2}\right)$$

where

$$v = \frac{2\pi d \Delta N}{\lambda}$$

d is the transducer length along the optical path, ΔN is the change in index of refraction of the medium produced by the passage of the sound

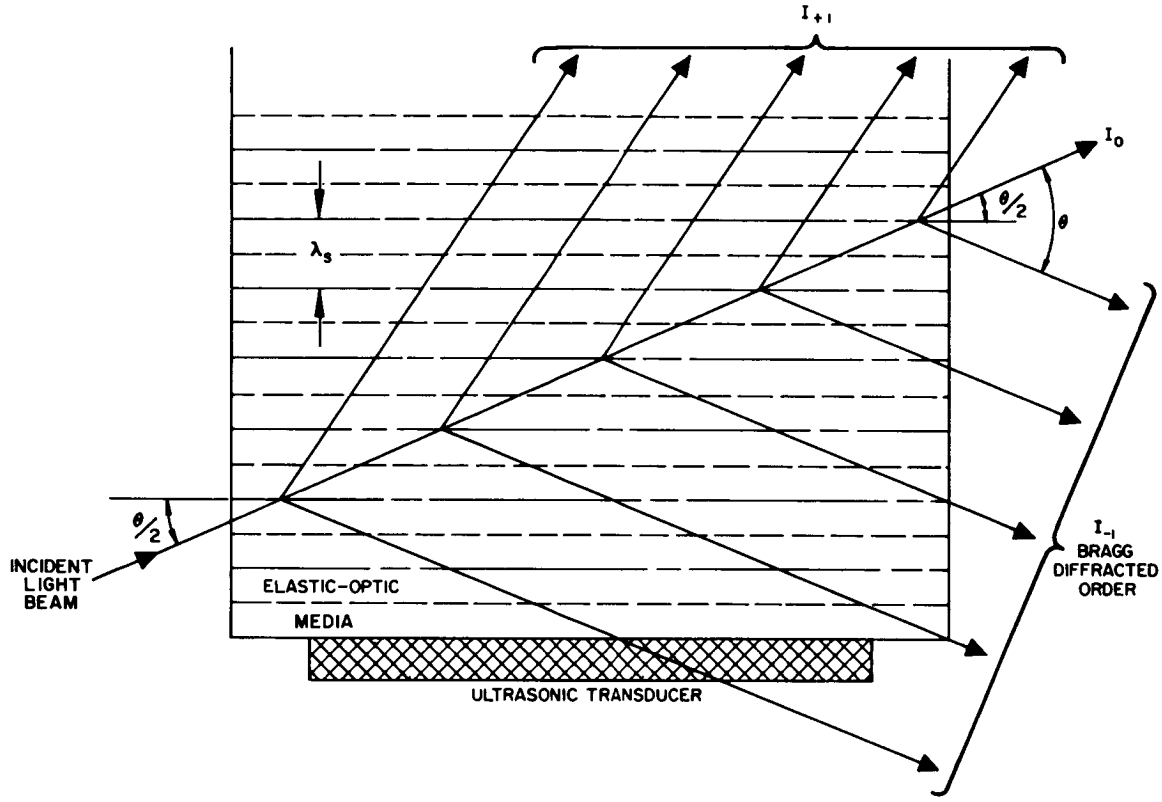


Figure 7-11. Ultrasonic diffraction geometry.

wave, and λ is the wavelength of light in vacuo. The change in index is related to the power in the ultrasonic wave by the equations

$$\Delta N = -\frac{1}{2} N_o^3 p_{ij} s_j$$

$$s_j = \sqrt{\frac{2S_a}{\rho v_s^3}}$$

$$S_a = \frac{TS_{rf} \times 10^7}{A}$$

Where S_a is the acoustic power density, ρ is the density of the medium, v_s is the velocity of sound of the medium, T is the electromechanical transduction efficiency of the transducer, A is the transducer area, and S_{rf} is the incident RF power. It has been shown experimentally that for sufficiently large values of v, the Raman-Nath parameter, approximately 95 percent of the energy incident on the modulator can be translated into the -1 diffraction order at Bragg angle. However, the choice of experimental conditions must be optimum and would seldom be encountered in practice. It is important to note that the modulation effect falls off with increasing wavelength so that for a modulator capable of producing 60 percent modulation at 0.6328 microns, approximately 1/300 of this or 0.2 percent modulation would be obtained at 10.6 microns.

7.3.3 Modulation Techniques

Frequency Translation

We have shown in the previous section that a traveling ultrasonic wave can diffract a light beam into an arbitrary number of diffraction

orders depending on the transducer length and the radiation geometry. The diffracted beams will be frequency shifted by an amount

$$\Delta f = f_o \pm nf_s$$

where f_o is the frequency of the light wave, f_s is the frequency of the sound wave, and n is the diffraction order. By adjusting the modulator for the Bragg condition, the intensity in all orders except -1 may be substantially reduced to the point where the device produces a single frequency shifted sideband I_{-1}. This mode of operation is entitled frequency translation and is useful in generating local oscillator beams for heterodyne detection systems. A detailed tabulation of the modulation efficiencies expected from a number of materials is presented in Section 7.3.5.

Intensity Modulation

If the traveling ultrasonic wave is reflected at a boundary of the modulation medium so that it returns to the transducer undeviated, a reflected or backward ultrasonic wave is generated, which, if unattenuated, is capable of producing a diffracted beam system spatially coincident with that produced by the forward wave and of equal intensity. Moreover, if the forward wave produces a positive frequency translation in a given order, the backward wave produces a negative one and vice versa. Thus the sideband spectra become double sidebands and act as suppressed carrier intensity modulated beams. The central or zero order beam becomes intensity modulated at twice the ultrasonic wave frequency since one now has an ultrasonic standing wave which diffracts energy away from the main beam twice per cycle. Because of the large values of mechanical Q usually encountered in ultrasonic resonators, the amplitudes of the forward and backward ultrasonic waves may reach values not obtainable in a traveling wave situation, and may therefore obtain large values of modulation index at low RF drive levels. However, this system is of limited usefulness where

large amplitude modulation bandwidths are desired because of the high mechanical Q of the system. One can, of course, use mechanical damping techniques to broaden the bandwidth of the modulator at the expense of modulation index and efficiency.

By proper choice of modulator dimensions and other parameters it is possible to produce a single diffraction maximum consisting of two frequency-translated optical waves. This system may prove valuable as a mechanism for locking of laser modes at infrared wavelengths, if the diffracted beam is fed back into the laser system.

If the optical faces of the modulator are made optically flat, highly parallel, and highly reflective, one can orient the crystal to obtain maximum transmission of light through the crystal at a selected set of optical wavelengths. This is the well-known Fabry-Perot effect. Similarly, if one passes sound waves through the crystal, one finds that the multi-reflected light wave interacts many times with the sound wave and in consequence the interaction cross section or modulation is stronger. This effect seems to be of limited usefulness in the visible region of the spectrum because of the change in modulator dimension when ultrasonic energy is dissipated in the crystal. At longer wavelengths this technique may be practicable because the effect of expansion is reduced, since the light wavelength is longer. With the use of good temperature regulation of the crystal, a Fabry-Perot acoustic modulator may be useful at 10.6 microns, where the dimensional tolerances will be about 20 times less than in the center of the visible spectrum.

Polarization Modulation

By focusing the optical beam to a diameter small compared to the ultrasonic wavelength it is possible to modulate the polarization of the optical beam, i.e., change the polarization from linear to elliptical, because of the birefringence induced in the modulation medium by the ultrasonic strain wave. The birefringence is simply calculated by taking the difference between the elasto-optic constants p_{11} and

p_{21} or p_{33} and p_{31} and multiplying by the appropriate variables to yield a value ΔN, the induced change in refractive index. Further discussion of the elasto-optic birefringence modulation will not be detailed here since the application of birefringence modulation is well described in the section on electro-optic modulators.

7.3.4 Properties of Ultrasonic Modulators

Choice of Acoustic Media

One sees from the foregoing treatment that when a traveling sound wave interacts with a traveling light wave in an elasto-optic medium (solid or liquid) one can obtain a frequency translated diffracted beam which may be used for system applications requiring frequency translation. At ultrasonic frequencies below 50 mc/sec one may use optically transparent liquids such as water, alcohol, tetrachloroethylene, carbon tetrachloride, etc., as the modulation medium, provided they are transmissive to the optical wave. At frequencies above 50 mc/sec the high ultrasonic attenuation in all liquids prevents one from obtaining satisfactory modulation indices and forces one to consider solids as a modulation medium, since they have intrinsically lower losses than liquids. In Section 7.3.5 a tabulation is presented which compares the relative performance of useful ultrasonic modulation media at one visible and two infrared wavelengths. A general result is that materials with large values of refractive index and low values of sound velocity are the most effective modulation media.

Anechoic Terminations

The development of single sideband frequency translators in elasto-optic media introduces a new problem, namely the absorption or termination of the ultrasonic wave. This operation is most conveniently done by shaping the modulator crystal so that the coherence of the ultrasonic wave is destroyed and it can no longer coherently diffract the light wave. Such terminations are termed anechoic chambers and are designed according to the same principles as optical black

bodies. However, the expense of cutting and shaping the single crystal is easily avoided by placing the crystal in contact with a pool of mercury which is contained in an anechoically shaped container. In practice it is found that mercury is a good acoustical impedance match for a large number of solids since the product of its density and sound velocity (acoustic impedance) is within 30 percent of that of many crystals. Thus the power reflection coefficient at a quartz-mercury or a quartz-silicon interface is of the order of 3 percent or less. Therefore, most of the ultrasonic energy passes into the mercury and is then rendered incoherent at the rough boundaries of the mercury container.

Transducer Bonding

Another problem common to solids but not encountered in liquids is the production of a satisfactory bond between the transducer and the modulator crystal. The bond should be lossless and of uniform thickness and it should be small compared to an ultrasonic wavelength. If all of these conditions are met then the transducer bandwidth and electrical impedance can be calculated or predicted in a straightforward manner. If they are not met, the bond can act as a mechanical transformer n quarter wavelengths long and thus preclude prediction of the frequency response and efficiency of the system.

In practice it has been difficult to produce ideal ultrasonic transducer bonds without a great deal of care for transducers operating in the 5- to 100-mc range. For frequencies above 100 mc a new transducer technique using evaporated thin films has some advantages not previously available; e. g., direct application of the transducer to the modulator without use of intermediate layers, fundamental frequency operation over a broad band, and fairly high transduction efficiencies.

Electrical Impedance Matching

Because of the high fundamental operation frequency, the transducers are very thin and thus present a high capacitance to the RF generator or amplifier. It is usually difficult to indictively compensate this capacitance over a broad band of frequencies without

using resistive damping and suffering an attendant increase in transduction loss. Theoretical calculations indicate that the use of a resonating inductance with a Q of 200 as an impedance matching transformer will yield an ultrasonic transduction loss of 10 dB, for a Quartz transducer radiating into quartz, germanium, or silicon. In practice transduction losses of the order of 12 dB are usually measured indicating the presence of other losses such as dielective loss, skin effect losses in wires, and dissipation in the ultrasonic bonds. If transducers with higher electromechanical coupling than quartz were used, e. g., ZnO, CdS, etc., one could obtain a factor 2 to 4 improvement in transduction efficiency. Thus materials such as CdS or ZnO are preferable for ultrasonic transduction since they enable the designer to obtain a larger electrical bandpass and at the same time a reduction in the transduction loss that is obtained with quartz. These materials are used for making thin film evaporated transducers since in single crystal form they are quite brittle and do not lend themselves to the production of large area transducers.

Transit Time Effects

In the Fabry-Perot modulator, as in the normal ultrasonic modulator, the transit time of the ultrasonic wave across the optical beam is a modulation bandwidth determining factor. For example, if the velocity of sound is 5×10^5 cm/sec and the laser beam diameter is 10^{-1} cm, then the transit time is approximately 0.2×10^{-6} sec or the bandwidth is confined to frequencies below 5 mc/sec. It is possible to use a long focal length lens to reduce the beam diameter to the diffraction limit, since the beam diameter at focus, a, is given by

$$a = \lambda \frac{f}{D}$$

where λ is the wavelength of light, ($\approx 10^{-5}$ - 10^{-3} cm), f is the focal length of the lens and D is the diameter of the laser beam incident on

the lens. Thus with the aid of a 10 cm focal length lens a light beam of diameter 10^{-1} cm at 6328 Å can be focused to a spot of dimension 6.3×10^{-2} cm corresponding to a transit time $\approx 10^{-7}$ sec or a bandwidth of 10 Mc/sec. Use of a shorter focal length lens can reduce the beam diameter even further if necessary to reduce the transit time of the ultrasonic wave across the light beam.

In effect, at the higher modulation frequencies,[17] the long transit time produces spatial modulation of the light wavefront and thus reduces the intensity of the modulation in a given diffraction order. Operating at frequencies low compared to the transit time will produce a percentage modulation in a given order approaching 100 percent, however operation at the frequency equal to the reciprocal of the transit time will reduce the modulation in a given order to about 30 percent of the maximum.

These considerations are usually important when one is concerned with highly efficient modulator systems. However, ultrasonic modulator systems in their present state of development are not highly efficient since the physical nature of the transduction system limits one to at least 6 dB transduction loss with the most advanced experimental transducers.

Summary and Tabulation

On the basis of the equations shown in Sections 7.3.1 and 7.3.2 the intensity in a modulated optical sideband produced by a traveling ultrasonic wave is given by

$$\frac{I_{-1}}{I_o} = \sin^2 \left[\frac{\pi d N_o^3}{2\lambda} \cdot p_{ij} \sqrt{\frac{2 TS_{rf} \times 10^7}{\rho v_s^3 A}} \right]$$

where $A = d \times h$, the transducer area, for small values of the argument or for optimum Bragg conditions, i.e., when the transducer is of

sufficient length that a single Bragg maximum is produced. This can be approximated by

$$\frac{I_{-1}}{I_o} = \frac{\pi^2}{2\lambda^2}\left(\frac{N_o^6 p_{ij}^2}{\rho v_s^3}\right)\left(\frac{d}{h}\right)\left(S_{rf}\, T \times 10^7\right)$$

Practical values of transduction efficiency T are of the order of 10 to 15 dB, corresponding to a power ratio of 0.1 to 0.03. Experimentally, it is found that the p_{ij} values range from 0.01 to 0.30 in most transparent solids, so that it is advantageous to select materials with as large a value of N_o^6 and as small a value of ρv_s^3 as possible.

Assuming a d over h ratio of 5:1, an RF power level of 20 watts, and a transduction efficiency of 0.03, the term in brackets takes a value of 3×10^7. The quantity $\pi^2/2\lambda^2$ takes a value of 1.23×10^9 at a wavelength of 0.633 μ, 4.3×10^7 at 3.39 μ and, 4.38×10^6 μ. An approximate table of values for I_{-1}/I_o for various materials follows (Table 7-3).

Table (7-3) indicates the potential of KRS-5 and CdS as candidates for producing 100 percent modulation at 6328 Å with 20 watts of RF drive power.

In the infrared region of the spectrum, Germanium and KRS-5 are the leading contenders for single sideband frequency translators but the production of high percentages of modulation at infrared wavelengths will require prohibitively high RF power levels (≈2kW). It is important to realize that a good transduction system can change the efficiency by a factor of 3 so that the possibility of 50 percent modulation at 10.6 microns exists.

Material	Relative Elastro-Optic Coefficients $\frac{N_o^6}{\rho v_s^3}(p_{ij})^2$			Modulation Ratio*** I_{-1}/I_o		
	0.63 μ	3.39 μ	10.6 μ	0.63 μ	3.39 μ	10.6 μ
Quartz	1.07×10^{-18}	2.66×10^{-19}		0.04	3.4×10^{-4}	
Cadmium Sulfide	24.4×10^{-18}	6.1×10^{-18}	4×10^{-18}	0.9	7.8×10^{-3}	5.3×10^{-4}
Zinc Oxide	10×10^{-18}	2.5×10^{-18} **	2×10^{-18} **	0.37	3.2×10^{-3}	2.6×10^{-4}
Silicon		8.9×10^{-18}	6×10^{-18} **		1.02×10^{-3}	7.9×10^{-4}
Germanium		5.3×10^{-17} **	4×10^{-17}		6.8×10^{-2}	5.3×10^{-3}
KRS5		2.7×10^{-17} **	2×10^{-17} **	≈ 1.0*	3.5×10^{-2}	2.6×10^{-3}
$LiNbO_3$				0.16		

*When the argument of $\sin^2(v/2) \approx (v/2)^2$ exceeds the range of validity of the approximation, one must return to $\sin^2(v/2)$ to avoid values of $I_{-1}/I_o > 1$.

**Most of the values of p_{ij} are estimated on the basis of the measured behavior of CdS in the visible and CdS and Si at 3.39μ. No actual measurements have been made at 10.6μ at this time.

***Values of the modulation ratio computed for 20 watts of electrical driving power applied to the transducer and 15 dB transduction loss.

Table 7-3. Elasto-optic performance parameters for selected crystals.

7.4 INTERNAL MODULATION

The term internal modulation is used to describe modulation schemes performed within the optical resonator.[19] This is also referred to as coupling modulation.

Three such schemes have been demonstrated. These will be discussed in the following sections. The common feature of these schemes is to transfer a fraction of the internal laser power from its original form, for which the resonator has a high Q (i.e., a high reflection), to a form in which it escapes readily from the resonator. This new form can be due to a modification of: (1) the polarization direction, (2) the direction of propagation, (3) the frequency.

The amount of power coupled in this manner can be made, as will be shown below, proportional to the (instantaneous) modulation voltage so that a square-law detector will recover the modulation signal.

The main advantage of these schemes over external modulation is that by acting on the internal laser energy, which may exceed the external power by more than two orders of magnitude, a given modulation power can give rise to proportionally higher absolute modulation of the output beam. This point will be discussed further below.

7.4.1 Internal Polarization Modulation

This form of internal modulation uses an electro-optic crystal inside the resonator. The crystal orientation and biasing with respect to the laser beam is identical to that which obtains when the same crystal is used externally to amplitude modulate the beam, i.e., the crystal couples a fraction

$$\frac{I}{I_o} = \sin\frac{2\Gamma}{2} \qquad (7\text{-}10)$$

from the original laser polarization to one in which the optical E vector is rotated by 90 percent degrees about the direction of propagation. Γ, the optical retardation, is given as in Section 7.2, by

$$\Gamma = \frac{2\pi\ell}{\lambda} N_o^{\,3} rE$$

for the transverse (to the direction of propagation) mode of operation. r is the appropriate electro-optic coefficient and E is the transverse electric field applied to the crystal. The other symbols are defined in Section 7.2.

In order to have a useful modulation scheme a way to couple out the modified radiation must be devised. This can be done by a polarizing prism (such as Rochon or Glan-Thompson prism) which separates the two incident polarizations, or, less efficiently, by Brewster windows which reflect a fraction, typically ~ 20 percent - 60 percent, of the second polarization.

As an example, consider the case when the applied field consists of a dc bias plus a time dependent modulation field, i.e.,

$$E = E_B + E_m(t)$$

letting $2\pi\ell/\lambda\, N_o^3\, r = a$, we have from (7-10)

$$\frac{I}{I_o} \approx \frac{a^2}{4}(E_B + E_m)^2$$

for

$$\Gamma = aE << 1$$

or when

$$E_m << E_B$$

$$\frac{I}{I_o} \cong \frac{a^2}{4}(E_B{}^2 + 2E_B E_m)$$

The radiation coupled out of the cavity is thus in the form

$$I \cong I_o \frac{a^2}{4}\left[E_B{}^2 + 2E_B E_m(t)\right]$$

and thus has a modulation index of $m = 2E_m/E_B$. A square law detector will yield an output

$$I_{detector} \propto I_o E_B E_m(t) \qquad (7\text{-}11)$$

thus recovering the original information signal $E_m(t)$.

The advantage of the internal scheme is now apparent. The radiation intensity inside the laser exceeds that outside by a factor equal to the inverse of the mirror transmittance. For 1 percent transmittance, to use a common value, $(I_o)int/(I_o)ext = 100$. It follows then from (7-11) that for a given modulation signal and the same E_B, the detector output is larger by 100.

A schematic diagram of a polarization modulation scheme is shown in Figure 7-12.

7.4.2 Other Methods of Internal Modulation

Another mode of internal modulation is based on the reflection of light from traveling sound waves.[20] A sound wave traveling in a medium (solid or liquid) placed inside the resonator intercepts the optical beam at the Bragg angle. The spatial modulation of the optical dielectric constant which is proportional to the local strain, acts as a diffraction grating and scatters a portion of the laser energy into the various diffraction orders. The amount of scatter intensity is proportional to the modulation acoustic power. In addition, the scattered beam is Doppler shifted in frequency by the acoustic frequency. One scattered beam is shifted upwards while a second scattered beam, emerging in a direction opposite to the first, is shifted downwards in frequency.

A third scheme consists of shifting a fraction of the internal energy in frequency and then extracting the frequency-shifted power by an etalon mirror which has a pass-band at this frequency.[21] The

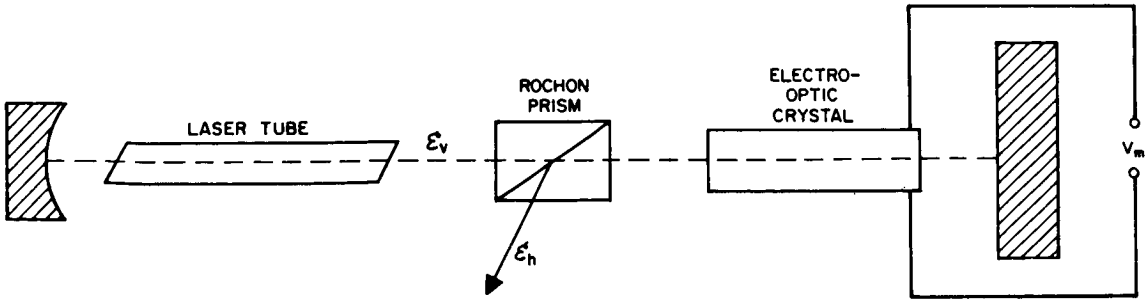


Figure 7-12. Laser oscillator with internal polarization modulation.

internal frequency shift is performed by low index phase modulation inside the laser resonator by an electro-optic crystal which has an induced optical axis parallel to the optical electric field. The amount of power coupled by this scheme and its dependence on the modulation voltages is similar to that described in Section 7.4.1. A schematic diagram of this scheme is shown in Figure 7-13.

This scheme suffers from the disadvantage of requiring an acoustic transducer which has to have a bandwidth comparable to that of the information.

Several techniques of frequency modulation and translation of optical laser beams have recently been reported which employ electro-optic modulators mounted within the Fabry-Perot resonant structure of the laser. Two such devices which perform quite different operations will be discussed here briefly.

The first system,[22] illustrated in Figure 7-14 is basically an SSBSC modulator and is related to the two-element scheme described in Section 7.2.3.2. The basic components of this system consist of a Nicol prism or its equivalent, an electro-optic modulator, a $\lambda/8$ wave plate, and a highly-reflective mirror. In Figure 7-14, the Brewster window of the laser serves as the Nicol prism.

Operation of this configuration is as follows: The vertically-polarized laser beam traverses the modulator whose induced axes of

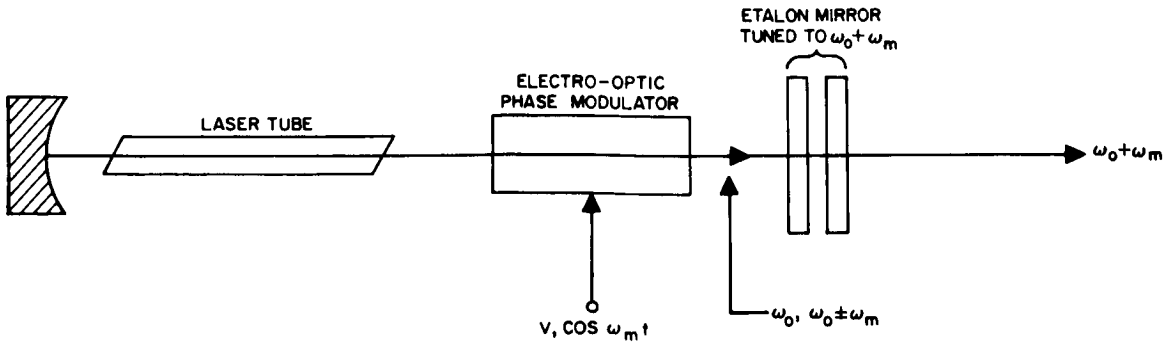


Figure 7-13. Laser with coupling modulation by internal frequency shifting.

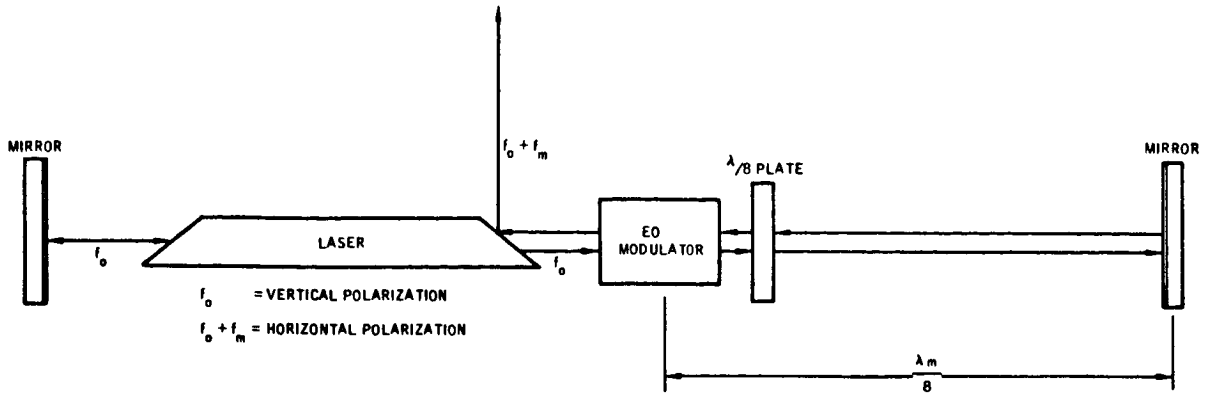


Figure 7-14. Internal SSBSC modulator.

birefringence lie at 45 degrees to the vertical. The beam then passes through the eighth-wave plate and is reflected from the end mirror, located a distance $\lambda_m/8$ from the center of the modulator, where λ_m is the free-space wavelength of the modulating signal. The return beam after traversing the $\lambda/8$ plate and modulator now contains a horizontal component shifted in frequency by f_m, and part of this is reflected out of the laser by the Brewster window. The double pass through the eighth-wave plate has the effect of rotating the modulator through 45 degrees for the return beam, and the round-trip photon transit time between modulator and mirror provides the 90-degree phase shift of the modulating signal.

This system can be used outside the laser as well, provided a real Nicol prism and additional reflector are used. In either configuration, the amplitude of the extracted beam is proportional to $J_1(\Gamma_o/\sqrt{2})$, as with the previous two-element modulator.

The second type of internal device[23] is simply an electro-optic modulator, oriented for pure phase modulation; i.e., the principal induced birefringent axis is parallel to the vertical polarization of the laser mode. When the modulation frequency is adjusted to be approximately but not exactly equal to the laser axial mode spacing, the laser mode can be made to oscillate with FM phases and nearly Bessel Function amplitudes. Thus the output is forced into the state of a

frequency-modulated optical signal which is swept over the entire Doppler-line-width at the modulation frequency.

This scheme promises to have valuable applications in some communication systems, but it appears to have a limited information bandwidth capability.

It should be pointed out that internal modulation schemes result in a compromise in laser design which may deleteriously effect the communication system. The mode selection problem becomes more severe and the laser output power is necessarily reduced because the reflection at the output mirror must be increased to compensate for the losses introduced into the cavity by the modulator.

7.5 MODULATOR PERFORMANCE-MODULATOR/BURDEN RELATIONSHIPS

It is premature at this time to make an assessment of the performance and burden characteristics of optical modulators which will have continuing value to the designer of laser systems for space communications and tracking. Very great progress has been made in the past few years in the development of optical modulation techniques. This work has demonstrated that all forms of modulation can be impressed on optical carriers in the band between 0.4 and 1.5 microns with laboratory apparatus. At wavelengths longer than 1.5 microns optical modulation technology is in a more primitive state wherein the available physical processes — primarily electro-acoustic — have provided modulation bandwidths far short of those believed to be needed in optical space communications systems. Research and development of acoustic modulation techniques is progressing at a moderate pace, but we believe it is too early to accurately assess the outcome of this work in terms of practicality for space system applications. Another possibility is that electro-optic crystals having suitable IR transmission will become available and provide the basis for extending the more advanced electro-optic modulation technology. Single-domain ferroelectric crystals, such as $LiNbO_3$ and $LiTaO_3$, which are transparent from 0.4 to nearly 5 microns and which possess a strong electro-optic effect, are now being successfully grown on a modest scale.

As implied in the above discussion, most of the development of optical modulation techniques has been carried out at the level of laboratory experimentation, with very little effort devoted to engineering for sophisticated applications. This laboratory type of work needs to be continued because the eventual achievement of acceptable performance versus burden characteristics of modulators for space optical systems probably can only be accomplished by the combined results of better materials and more efficient interaction structures, followed by intensive engineering development.

Table 7-4 is presented merely as a preliminary data sheet to be used with these precautionary remarks in mind. Some of the data presented are ambiguous because the same ground rules are not used in defining terms. For example, the 3dB bandwidths in some cases refer to the inherent bandwidth of the modulator interaction structure, while in other cases the value defines the operating performance of a specific experimental modulator when it is driven by a specific driver. In some cases the usable bandwidth depends on the combined characteristics of the modulator and the driver. The weights and sizes shown for some of the modulators are values obtained when negligible emphasis was placed on these characteristics.

Parameter	TM_{oo} Mode Cavity Modulator	TW Strip-Line Intensity Modulator	Zig-Zag Intensity Modulator	Strip Line Polarization Modulator (NASA Contract)	Multi-Element Intensity Modulator	Resonant SSBSC Modulator	TW Strip-Line SSB Modulator	TW Strip-Line Intensity Modulator
Development Status	Built at BTL	Built at Sylvania	Dev. at BTL	Built at Hughes Aircraft Company	Built at Hughes Aircraft Company	Built at Hughes Aircraft Company	Built at Hughes Aircraft Company	Designed at Sylvania
Material	KDP	KDP	CuCl	KDP	KDP	KDP	KDP	KDA
Crystal Dimensions	1 cm long × 2.5mm dia.	100 cm long × 2mm wide	6.7 cm long × 1mm sq	50 × 0.4 × 0.4 cm	16 each 1/4" × 1/4" × 1/2"	2 each 1 1/2" long × 1/2" dia.	80 × 0.4 × 0.4 cm	18 × 0.2 × 0.2 cm
Optical Attenuation	0.1 dB	6 dB	2.6 dB	1.5 dB	0.3 dB	1.0 dB	2.0 dB	Unknown
Useful Optical Range	0.4 - 1.5 μ	0.4 - 1.5 μ	0.4 - 20 μ	0.4 - 1.5 μ	0.4 - 1.5 μ	0.4 - 1.5 μ	0.4 - 1.5 μ	0.4 - 1.5 μ
Modulation Frequency	3 Gc	Baseband	Baseband	Baseband	Baseband	850 Mc	200 Mc	Baseband
3 dB Bandwidth	4 Mc	3 Gc	> 10 Gc	30 Mc*	10 Mc*	5 Mc	100 Mc	3 Gc
Modulating Power	1.5 W	12 W	1.6 W	20 W	12 W	3 W	10 W	5 W
Modulation Index	0.13	~1.0	0.5	0.5	0.5	0.01	0.3	0.3
Extinction Ratio	Unknown	Unknown	Unknown	8:1	250:1	NA	NA	Unknown
Weight	Unknown	Unknown	Unknown	20 lbs	2 lbs	15 lbs	25 lbs	Unknown
Size	Unknown	Unknown	Unknown	144 in^3	30 in^3	850 in^3	216 in^3	Unknown

*3dB Bandwidth limited by modulator driver and not by the electro-optic structure.

Table 7-4. Characteristics of some electro-optic modulators - June 1966.

7.6 NOMENCLATURE

Γ	relative phase shift between principal components of a polarized light beam induced by a birefringent element; e.g., an optical modulator.
Γ_o	peak value of Γ
I	output intensity of a modulated light beam.
I_o	beam intensity at input to modulator
N_x, N_y, N_z	principal refractive indices of an optical material for light polarized in the x, y, z direction, respectively.
N_o	ordinary refractive index of a birefringent material
N_e	extraordinary refractive index of a birefringent material
λ	optical wavelength
r_{63}, r_{41}, etc.	linear electro-optic (Pockels) coefficients
g_{11}, g_{12}, etc.	quadratic electro-optic coefficients
E	electric field, DC or AC
E_m	electric field, specifically an applied modulation field
V	voltage, DC or AC
V_m	voltage, specifically an applied modulation voltage
$\mathcal{E}$	electric field of an optical beam
$\mathcal{E}_x$, $\mathcal{E}_y$	x, y components of an optical electric field
$\mathcal{E}_h$, $\mathcal{E}_v$	horizontal, vertical components of an optical electric field
ω	2π times an optical frequency
ω_m	2π times a modulation signal frequency
m	modulation index
m_i	effective modulation index of the i^{th} harmonic of a modulation signal

ϵ	permittivity or inductive capacity of a dielectic medium
ϵ'	real part of ϵ
ϵ''	imaginary part of ϵ
k	temperature coefficient of natural birefringence
B	the Kerr constant
ϕ_h, ϕ_v	total phase retardation of the horizontal, vertical components of an optical beam
®	temperature in degrees C (or K)
tan δ	loss tangent of a dielectric material
M	general symbol for a 2 by 2 matrix operator
M_i	matric operator representing the i^{th} member of a series of optical elements
θ	an angle, usually representing relative orientation of two coordinate systems
V_x, V_y	x, y components of an applied voltage
V_o	amplitude or peak value of an applied voltage
$J_n(x)$	n^{th} order Bessel function of an argument X
B_{ij}, B_{jk}, etc.	the dielectic impermeability tensor or indicatrix, a dimensionless quantity
P_{ijk}	the elastro-optic coefficient of the medium, dimensionless
ηk	the strain induced in the elasto-optic medium, dimensionless
κ_{ij}, etc.	the dielectric constant of the medium, a dimensionless quantity
s_j	a strain component of η_k for a principal crystallographic direction
λ_L	the free space wavelength of the optical beam
λ_s	the sound wavelength in the diffracting medium

v	the dimensionless Raman-Nath parameter and is equal to $2\pi d\Delta N/\lambda$
S_a	the acoustic power density in ergs per second per square centimeter
S_{rf}	the total radio frequency or microwave frequency power input to the transducer
ρ	the density of the medium
v_s	the velocity of sound in the medium in units of centimeters per second
T	the transduction efficiency, dimensionless
A	the transducer area
f_o	the frequency of the optical wave
f_s	the frequency of the ultrasonic wave
n	the diffraction order number
I_{-1}, I_o, I_{+1}	the intensities of the n = -1, n = 0, and n = +1 order diffracted beams
d	the transducer length along the light path in centimeters
h	the transducer height in centimeters
f	the focal length of a lens
D	a laser beam diameter

7.7 REFERENCES

1. I. P. Kaminow, Phys. Rev. Letters 6, 528 (1961).

2. C. J. Peters, NEREM Record, 1964, p. 70.

3. A. N. Winchell, "Microscopic Characters of Artificial Minerals", (John Wiley & Sons, New York, 1931).

4. A. R. vonHippel, Dielectric Materials and Applications (John Wiley and Sons, New York, 1954), p. 301.

5. A ruby laser Q-switch employing the same principle of operation has been described by J. L. Wentz, Proc. IEEE 52, 716 (1964).

6. S. E. Harris and A. E. Siegman, Dept. Electrical Engineering, Stanford University, internal report.

7. C. F. Buhrer, L. R. Bloom, and D. H. Baird, Applied Optics 2, 839 (1963).

8. C. F. Buhrer, V. J. Fowler, and L. R. Bloom, Proc. IRE 50, 1827 (1962).

9. C. F. Buhrer, Proc. IEEE 52, 969 (1964).

10. R. Targ, G. A. Massey, and S. E. Harris, Sylvania Electronic Systems, Mountain View, California, internal report.

11. R. Targ, Proc. IEEE 52, 303 (1964).

12. C. J. Peters, Proc. IEEE 51, 147 (1963).

13. I. P. Kaminow and J. Liu, Proc. IEEE 51, 132 (1963).

14. J. F. Nye, Physical Properties of Crystals, New York; Oxford, 1960.

15. R. S. Krishnan, Progress in Crystal Physics, Madras; Central Arts Press, S. Viswanathan, 1958.

16. C. F. Quate et. al., "Interaction of Light and Microwave Sound", Proc. IEEE, Vol. 53, pp 1604, 1623, (1965).

17. H. V. Hance and J. K. Parks, "Wide Band Modulation of a Laser Beam using Bragg Angle Diffraction by Amplitude Modulated Ultrasonic Waves", J. Acoust. Soc. Am. Vol. 38, pp 14-23, July 1965.

18. M. G. Cohn & E. I. Gordon, "Acoustic Beam Probing using Optical Techniques", BSTJ, Vol. 44, pp 693-721, April 1965.

19. K. Gürs and R. Müller, Phys. Letters 5 (1963) p. 129.

20. A. E. Siegman et. al., Appl. Phys. Letters 5, 1 (1964).

21. Don G. Peterson and Amnon Yariv, Appl. Phys. Letters 5, 184 (1964).

22. R. Targ, G. A. Massey, and S. E. Harris, Sylvania Electronic Systems, Mountain View, California, internal report.

23. S. E. Harris and R. Targ, Appl. Phys. Letters 5, 202 (1964).

8.0 ACQUISITION AND TRACKING

8.1 INTRODUCTION

8.1.1 Study Status

The study effort of acquisition and tracking has been directed towards the following goals: (1) gaining a thorough understanding of the basic constraints within which the system must operate and (2) examining the effects of these constraints on some general configurations. As a result, detailed additional burden relationships, beyond those given in the Phase I Final Report, have not been derived. Also, the documentation of the acquisition and tracking state-of-the-art has proceeded at a low level.

The material contained in this section is organized as follows: Subsection 8.2 is devoted to general theoretical considerations affecting both the vehicle and earth-based systems. In addition, the DSV acquisition and tracking systems are considered, including a detailed treatment of the sources and effects of noise and angular errors. This subsection contains a discussion of the tracking problem with respect to the earth station and includes a brief discussion of the problems associated with a near-earth relay satellite.

Subsection 8.3 has preliminary performance data relevant to the state-of-the-art in star trackers and sun sensors. Subsection 8.4 is devoted to burden relationships of the acquisition and tracking system. Subsection 8.5 gives the nomenclature, and subsection 8.6 is a list of references.

8.1.2 Summary of the Study during the Last Quarter

The two factors which heavily influence the design of the deep space acquisition and tracking system are: (1) the long range with its associated transit time of an electromagnetic signal, and (2) the extremely narrow beamwidths required for high data rate communications.

At ranges greater than 10^7 n. mi., the signal transit time becomes on the order of minutes; indeed, during the terminal phase of a typical

Mars mission, the round trip transit time is approximately 15 to 20 minutes. Under such circumstances, the operation of the Deep Space Vehicle (DSV) transmitter - receiver system in a closed loop mode with the earth is greatly complicated, if not precluded. Therefore, the tracking and pointing error signals, which are the inputs to the attitude control systems must be generated internally, i. e., without feedback from the earth. When this restriction is combined with the extreme pointing accuracies dictated by beamwidth in the range of 1 to 50 μrad, one sees that the tracking problem is unlike any previously encountered. In addition for relative tangential velocities greater than zero, the aberration of light becomes important. In fact, if the tangential velocity of the receiver relative to the transmitting station is such that the receiver will move during the electromagnetic transit time, a distance greater than that projected by the transmitting beam, it will be necessary for the transmitter to lead the receiver by some angle. Thus, very narrow beamwidths impose a prediction tracking requirement on both the DSV and earth station transmitters.

Atmospheric perturbations must also be considered. In an earth based station, the perturbations to the transmitted and received radiation, due to the effects of the atmosphere, must be included in the design study. Attenuation due to scattering and absorption degrades the signal strength with an attendant degradation of the discrimination ability.

There are several other atmospheric effects which affect the acquisition and tracking. These are associated with anomalies due to turbulence and other atmospheric inhomogenieties and may be critical to the design study. Beam steering, scintillation, and quivering are among the important effects. These are discussed in some detail in Section 12.0, "Background Radiation and Atmospheric Attenuation".

In discussing the acquisition and tracking problem as it specifically relates to the DSV, it is first necessary to investigate the relative advantages and disadvantages of manned versus unmanned vehicles. This is done in the following subsection. Unmanned missions, of course, are not subject to man-induced motions and the burden they impose on the attitude control system. On the other hand, the versatility of an

astronaut and his ability to make accurate measurements with comparatively crude instruments could eliminate many of the problems associated with acquisition, tracking, and lead angle determination.

In this report, the primary emphasis is on the unmanned deep space mission, with the receiving terminal located either on the earth or on an earth satellite. Future reports will investigate manned missions and communication between two deep space vehicles.

In general, the DSV acquisition and tracking system differs from the earth-station system in three areas: (1) the higher data rates used in the DSV to earth communication link, (2) DSV available size and power restrictions; and (3) DSV environment restrictions. Areas (1) and (2) combine to dictate a narrow beam. The absence of an atmosphere aids in making such a narrow beam possible.

Boresight considerations and the requirement of adjustable lead angle indicate that the transmitter and receiver should use the same optical system.

The acquisition of the earth beacon by the DSV will procede in three steps: (1) approximate orientation of the receiver axis along the proper LOS; (2) search, and (3) transfer from search mode to the tracking mode. The use of a sun sensor and a star tracker with the predicted LOS as given by the DSIF will serve to accomplish the initial orientation of the receiver axis to within ± 40 μrad (rms).

Transfer from the wide-angle mode to the narrow-mode may be accomplished in several discrete steps or, possibly, continuously with a zoom lens. These are considered in more detail late in this section.

Once acquisition is complete, the LOS must be tracked and the telescope pointed to within a fraction of a beamwidth. Such accurate pointing almost certainly requires a body-fixed telescope for narrow beams (e. g., 1 to 50 μrad). The most promising method of control then appears to be a system of momentum storage device; in particular, inertial flywheels.

The pointing system will consist of: (1) the spacecraft attitude control system for coarse pointing; (2) a two-axis system to orient the entire telescope or, alternatively, an associated mirror system, and

(3) an internal beam deflector to produce the variable lead angle. Again, these are considered in more detail later in the section.

Stabilization of the vehicle is obviously integrally related to its pointing (and tracking) capability. The vehicle must be designed carefully to minimize external and internal torquing effects and sensing techniques, accurate to a fraction of a μrad/sec, must be developed.

Reaction wheels show the most promise of fulfilling requirements of fine vehicle stabilization. They will be investigated in more detail and the state of the art will be documented during the next quarter.

8.1.3 Study Plan for Next Quarter

The study next quarter will consist of a broadening of the theory section in this report, with special emphasis on the deep space vehicle, the earth station, and earth satellite terminals for unmanned missions. Vehicle attitude control will be examined in more detail to determine the tradeoff parameters for weight, size, propellant consumption, and reliability in terms of control limit cycle magnitudes and types and accuracies of inertial measuring devices required.

Vehicle attitude parameters thus obtained will provide the details for an investigation of the types and sizes of gimbals required for the communication system. A program will be initiated for simulation of logical servo systems to determine the major servo system gains and the compensation networks which will be required. Additional emphasis will be placed on the specific error anticipated in each type of system investigated. Recommendations for gimbal placement and alignment within the vehicle will be made to minimize the effects of vehicle rates.

A section on the probability of acquisition, and probability of loss of a tracking beacon, in the presence of background noise will be included to consider both the earth and stellar backgrounds for the DSV and the earth satellite receivers, and the stellar and atmospheric backgrounds for the earth station.

Performance studies will be continued for star and sun sensors and momentum storage devices.

8.2 GENERAL THEORETICAL CONSIDERATIONS

8.2.1 Geometrical Considerations

Acquisition and tracking is concerned with maintaining a line-of-sight; yet line-of-sight is an ambiguous term for a deep space mission. Actually there are two line-of-sight position vectors, neither of which physically terminates at the earth or DSV. For this report the LOS vectors are defined in Figure 8-1. These vectors will be called the earth-to-deep space vehicle (E-DSV) and the deep space vehicle-to-earth (DSV-E) vectors, respectively. Actually the DSV-E line-of-sight vector is directed to a point in space where the earth was when the received light left the earth on its way to the DSV. Similarly, the E-DSV line-of-sight vector is directed to a point in space where the DSV was when the light left the DSV. Clearly both the earth's receiver and DSV's receiver must be directed along the E-DSV and the DSV-E line-of-sights, respectively.

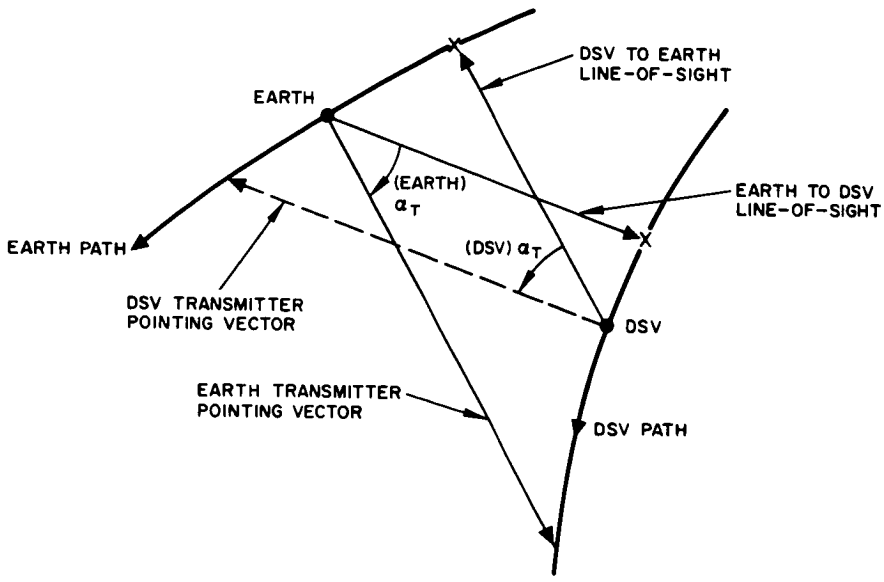


Figure 8-1. Earth and DSV line-of-sight geometry. DSV-to-Earth LOS shown at instant a signal received from earth. Earth-to-DSV LOS shown as a signal is received from DSV.

The transmitters must each be directed to points in space where the respective receivers will be when the transmitted light gets there. Thus a lead angle α_T is required for both the earth transitter and DSV transmitter. If atmospheric effects are neglected, $\alpha_{T(Earth)} = \alpha_{T(DSV)}$. In this case the earth transmitter pointing vector will be parallel to the DSV-E LOS. Likewise the DSV transmitter will be parallel with the E-DSV LOS. Atmospheric effects will cause the transmitted beams to be distorted and bent. This effect is not self compensating since the effect is different for a far field wave front than for a near field front. Thus $\alpha_{T(Earth)} \neq \alpha_{T(DSV)}$ in general. This effect must be accounted for if it is desired to update the DSV lead angle from knowledge of the earth's lead angle.

It can be shown that the lead angle is a function of relative tangential velocity where

$$\alpha_T = \frac{2V_T}{C}$$

For interplanetary missions typical relative velocities are on the order of 5 to 10 miles/sec. Thus values of α_T are on the order of 50 to 100 microradians. For a beam width greater than 100 micro-radians, as used in long wavelength transmissions, the concept of lead angle is not important and the LOS vectors can be considered parallel.

In this study we will be considering beam widths of the order of one microradian. Thus the lead angle and LOS distortions are of critical importance.

8.2.2 Atmospheric Effects

The main effects of the earth's atmosphere on the acquisition and tracking problem can be divided into four categories:

1. Attenuation due to scattering and absorption of the incoming or outgoing radiation by the various atmospheric constituents;
2. Beam spreading due to turbulence;
3. Beam steering or deflection due to refraction;
4. Scintillation (or twinkling) - variations in received and transmitted power produced by inhomogenities and turbulence.

Section 12.0 gives a reasonably complete discussion of attenuation effects. On a clear day, except for occasional cloud obscuration, less then 10 percent of the transmitted or received power is lost. Natually, as the visibility decreases the power loss increases. Under severe conditions - e. g., heavy fog - the power transmitted through the atmosphere approaches zero. However, for a suitable choice of transmitting and receiving stations, the effects of atmospheric attenuation should not be unduly restrictive.

There is a scarcity of available quantitative data on the effects of turbulence on beam divergence. Qualitatively it should be noted that there is a fundamental difference between the fluctuations in signal level during transmission and those during reception. During reception, although diffraction at the spacecraft spreads the beam over a large portion of the earth, all the energy incident on the receiver aperture can be detected if a sufficiently large field stop (detector) is used. However, during transmission, only that portion of the beam that leaves the atmosphere in the direction of the spacecraft is used. There is only a slight spreading of the beam during passage through the atmosphere, but angular or phase disturbances are created because the plane wave front has been distorted. These disturbances may result in a large spreading of the beam after subsequent propagation. Angular divergence here, perhaps not yet affecting beam diameter because of the large initial diameter, will ultimately be the determining factor in beamspread. In antenna phraseology, the top of the atmosphere is still in the near-field region of an optical transmitter. The basic difference between transmission and reception can be summarized as follows: in transmission, cumulative phase fluctuations (which cause angular divergences) are important; however, in reception, only the cumulative amplitude fluctuations (produced by phase fluctuations near the top of the atmosphere) are significant.

Beam steering arises from time-dependent atmospheric inhomogenities - e. g., turbulence - and from time-independent inhomogenities, such as steady state density gradients. It introduces an inherent angular error in specifying the true direction of the

earth-to-DSV line of sight vector. This in turn means that the earth-based transmitter must broadcast over a correspondingly wider angle to insure that the DSV is illuminated.

The amount of beam deflection which can be expected depends directly on the strength of the turbulence. As the turbulence goes from weak to strong, beam steering angles typically vary from ±1 to ±15μrad (rms). Very strong turbulence can produce deflections on the order of ±50 μrad (rms). In addition, the apparent direction of the LOS will vary in time due to quivering at frequencies on the order of a hundred cps or less. These (relatively) rapid angular variations can amount to the same order of magnitude as in the slowly varying part mentioned above.

Note that the upper limits quoted refer to daytime activity. At night the turbulence decreases sharply due to the decrease in thermal gradients.

Scintillation introduces random angle noise into the angle trackers with a power spectrum that varies as $f^{-2/3}$ near dc and has a high frequency cutoff determined by wind velocity. For large aperture telescopes this cutoff will usually occur at only a few cps.

8.2.3 Manned Versus Unmanned Vehicles

The advantages of a manned mission from an acquisition and tracking point of view are manifold. In conjunction with an on-board computer, man can make very accurate course and attitude correction and navigational sightings. Thus man is better capable of determining the DSV's course, attitude, and position than is the earth. Furthermore the astronaut can align an inertial platform reference device and make corrections for drift. Two other important advantages of manned flight are the facts that continuous tracking and attitude control are not required as are for the unmanned mission. A manned mission has the added advantage of accurate lead angle alignment, eliminating the errors associated with the servo device necessary to the unmanned system. The lead angle may be read out with optical techniques and adjusted when desired. An additional advantage to a manned mission is the

ability of the man to perform routine maintenance and repairs; an unmanned mission would require completely redundant systems for the equivalent reliability.

The primary disadvantage of a manned mission is the effect of man motion on the spacecraft, with resulting additional stringent restrictions on the tracking and pointing control system in order to avoid degradation of the tracking and pointing accuracy. A second disadvantage is the requirement for larger gimbal angles capabilities. A man in the spacecraft will undoubtly require freedom of spacecraft orientation for observation during communication, especially for planetary observation.

For the manned mission several disadvantages may be circumvented and several advantages of the unmanned system acquired if a separate "satellite" vehicle were used. Such a vehicle would be maintained by the astronauts if it could be repaired, the lead angle accurately set, and a platform could be used and aligned. When communication is desired, the satellite vehicle could be separated from the mother vehicle, and maneuvered in space by remote control. The satellite vehicle might be tethered to the mother vehicle or rendezvous techniques used for retrieval when repairs, alignments, course changes, etc., are to be made.

An unmanned mission will have the restraints of continuous tracking and continuous attitude control. An inertial platform reference device is not feasible for this system since drift correction would be very difficult. In place of a platform, star trackers and sun sensors could be used for attitude reference and control.

The general block diagrams for the unmanned mission and for the satellite vehicle of the manned mission are shown in Figure 8-2 and 8-3 respectively. For both missions, the electromagnetic transit delay is involved twice in the DSV-E-DSV communication loop. Cross-coupling of the dynamic relationships is indicated by the dotted lines. An inertial platform and rate gyro stabilization is used on the satellite vehicle for the manned missions. Due to offset and drift problems, star trackers (or similar devices) are used for rate and attitude

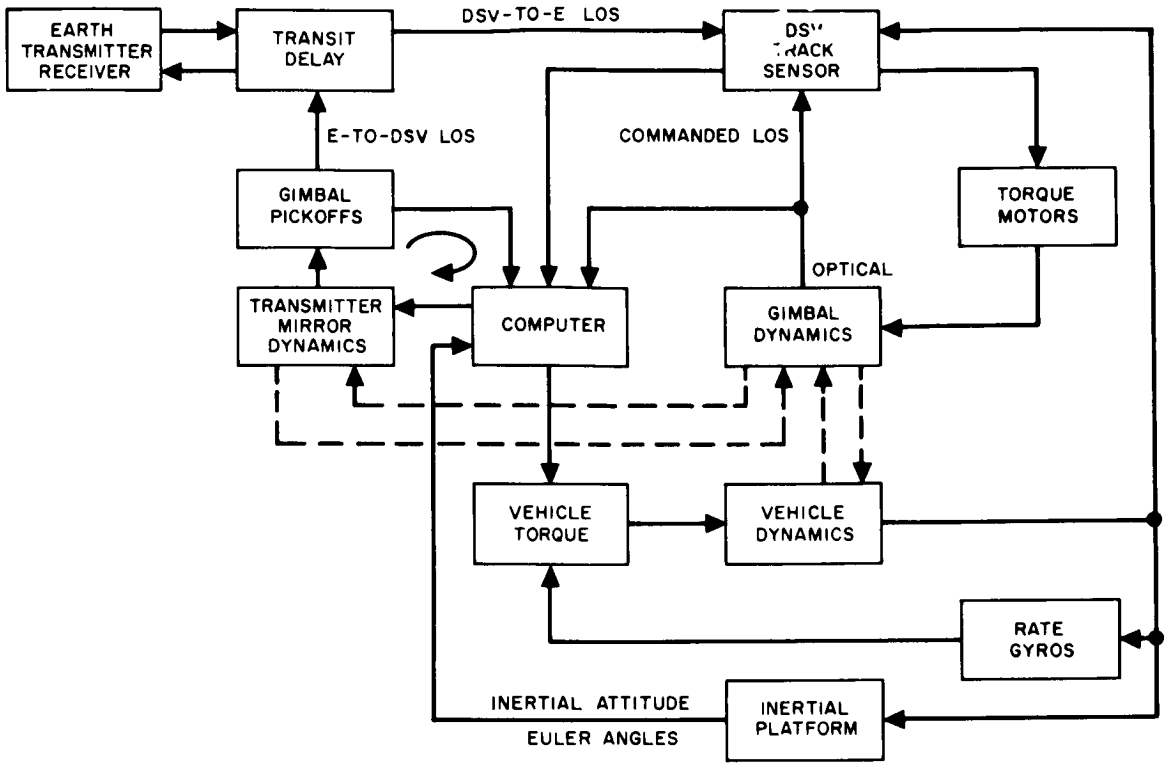


Figure 8-2. Deep space vehicle communication control system block diagram for a manned mission.

stabilization for the unmanned missions. The computer shown is intended for the complex navigational and lead angle computations. However much of this burden may be removed by an earth based computer utilizing the predictive capabilities of the DSIF.

8.2.4 Deep-Space Vehicle Station

The station at the DSV end of the communications link and the earth station are similar, in that both consist of input and output devices, a transmitter, a receiver, a tracker, and signal processing electronics. Three important differences are the higher data rate assumed to be required for DSV to earth transmission, available size and power in the DSV, and environment of the DSV. As a result of the first two differences, a narrower beam is required for the down link and as a result of the absence of an atmosphere at the DSV, such a narrower

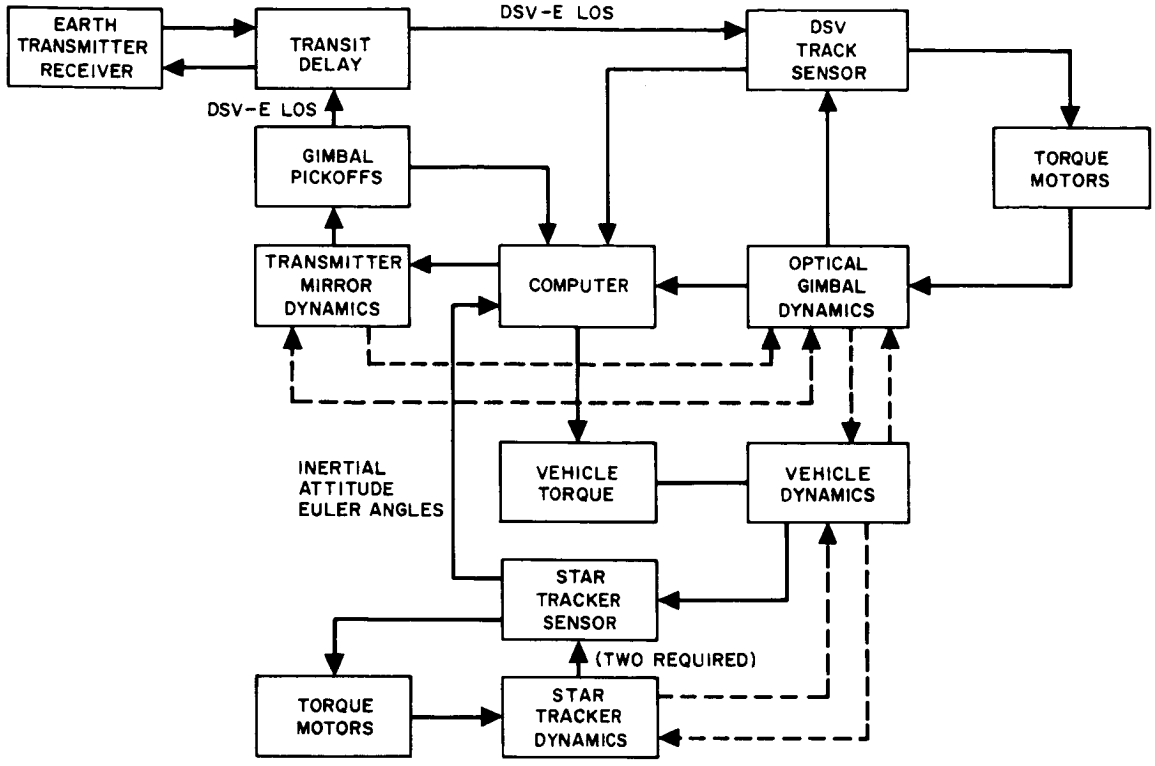


Figure 8-3. Deep space vehicle communication control system block diagram for an unmanned mission.

transmit beam is possible. The details of making the use of a very narrow transmitter beam (in the range of 1 to 50 microradians) feasible are the major tasks of the optical and mechanical design of the spacecraft telescope. Accurately pointing and controlling such a narrow beam are the major tasks of the acquisition and tracking control system.

In order to insure boresight integrity between the DSV transmitter and receiver, they should use the same primary optical system. If the transmitter and receiver use different wave lengths, separation of the transmitted and received signals can be accomplished spectrally, and the full aperture can be used for each. The exact spectral separation is an optical problem and will not be discussed in this section.

8.2.5 Earth Acquisition by the DSV

In order for the DSV to acquire the earth, an adequate earth beacon is assumed. Thus, the problem of acquisition reduces to one of orienting the DSV receiver field-of-view such that the earth beacon falls within it.

The sequence of events for the acquisition and stabilization is described in the following paragraphs. Block diagrams of the vehicle control system are shown in Figures 8-4 and 8-5.

The spacecraft must first be oriented such that the earth falls within a solid angle specified by the system gimbal limits. This is accomplished as follows: any large residual angular rates are first eliminated through operation of gas jets controlled by signals from a set of three rate gyros, one for each principal axis. The vehicle will then be oriented to point the telescope directly away from the sun by means of the gas jets and two-axis sun sensors. Roll rate about the telescope axis will be reduced to the limit cycle and held. Attitude signals for the pitch and yaw axes will be generated by the sun sensors,

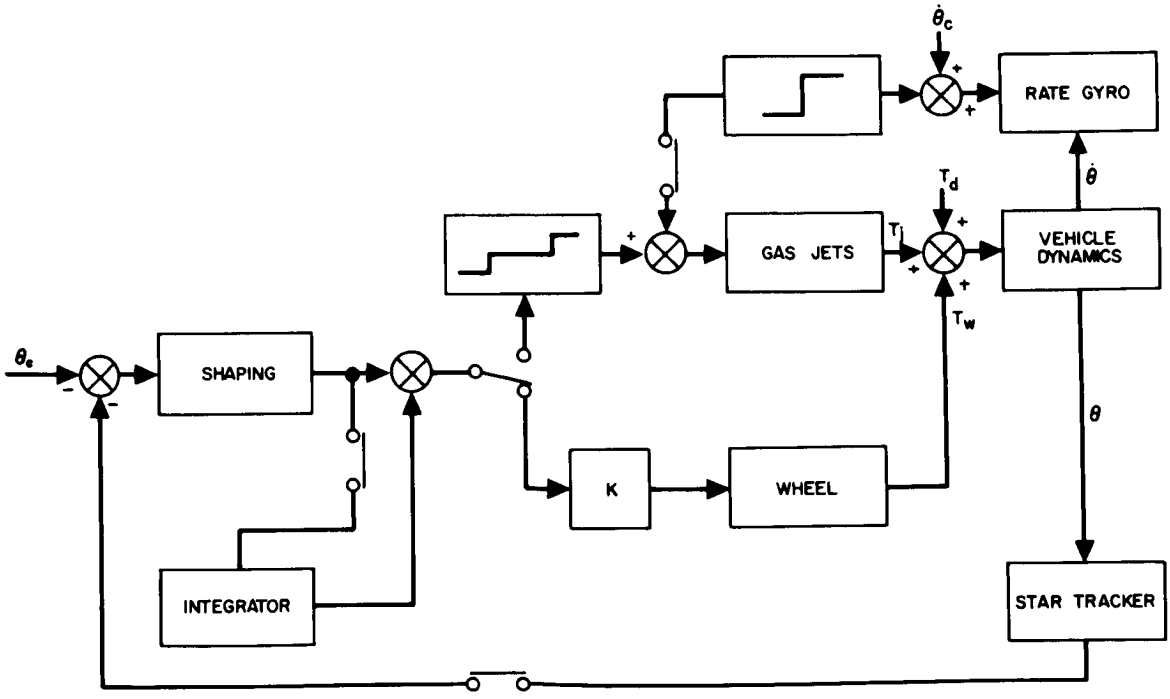


Figure 8-4. Unmanned DSV roll control system.

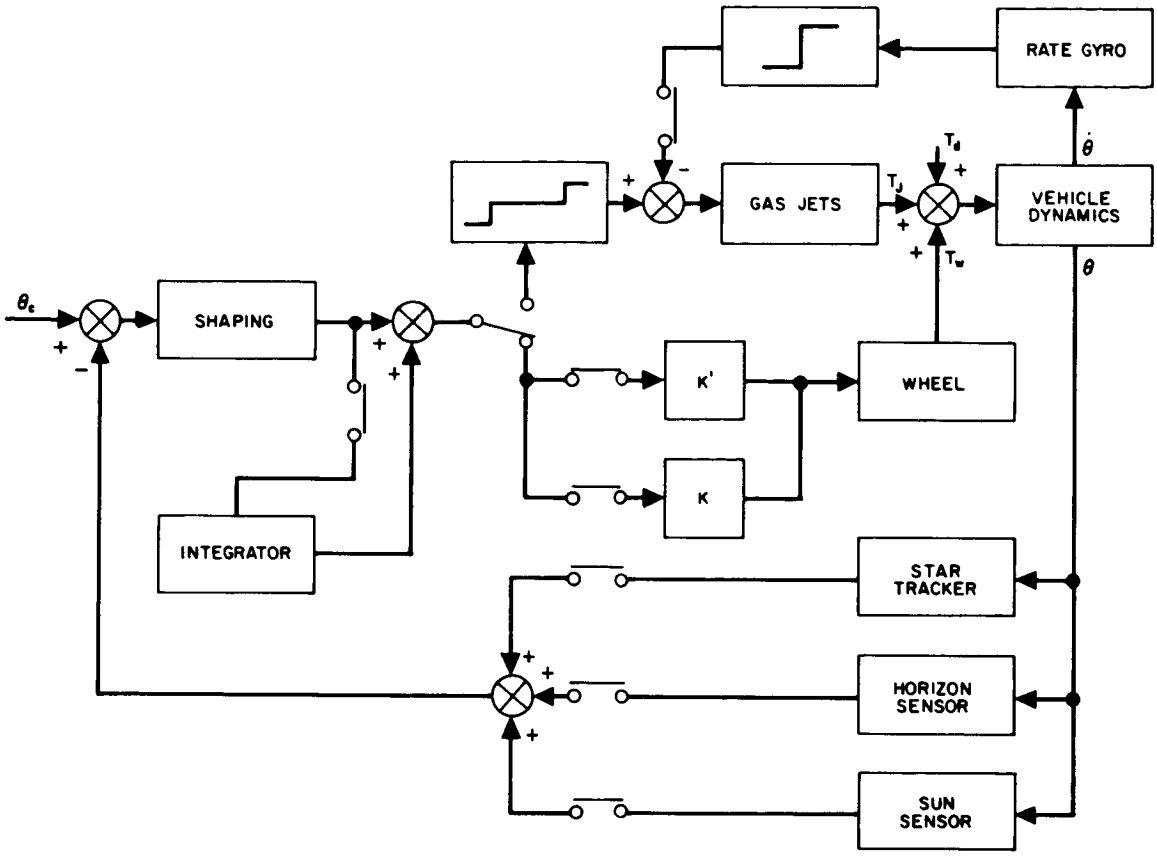


Figure 8-5. Unmanned DSV pitch or yaw control system.

which are nulled when the telescope looks away from the sun. Next is the acquisition of the roll reference star (Canopus or a similar star) by the roll star tracker. This will be accomplished by a command roll rate about the sun line of sight (via gas jets) until the reference star enters the field of view. The generated star tracker error signal is then switched into the roll loop and acquisition is accomplished in the same manner as in the pitch-yaw axes. Care must be taken to select the vehicle search rate small enough that the reference star may be acquired before the star passes through the field of view. Acquisition of the star line-of-sight is now completed and an inertial reference has been established. Command signals are now given for a preprogrammed attitude maneuver and the spacecraft is rotated to point the telescope in the vicinity of the earth. (At near earth ranges, the horizon sensor can

be used at this point.) When limit cycle operation has been achieved, the system is ready for the second phase of the acquisition.

The volume of space to be searched is scanned by the receiver using the telescope in a wide angle field of view mode. The scan optics of the telescope start at the edge of the solid angle and search toward the telescope axis until the earth beacon is detected. When the beacon is detected, the error signals are switched into the gimbal drives of the telescope and the telescope axis is oriented along the DSV-E line-of-sight. The system is now in a coarse error detection mode.

The space vehicle's communication system is now switched from the wide-angle search to the narrow angle track mode. It may be necessary to do this in several steps if the wide angle tracking error is greater than the telescope's narrow angle field-of-view. A zoom lens could be used to advantage here. If the "zoom" were slower than the response time of the gimbal control system, the pointing error could be corrected during the beam narrowing period. The system is now in the fine error detection or tracking mode of operation.

The sensor can now be switched out of the vehicle pitch-yaw control channels, and be replaced by the tracking error signals. The star tracker must be retained however, since the two-axis error information cannot be transformed into three-axis control error commands. This marks the completion of the acquisition sequence and tracking control mode begins.

The initial step in the acquisition phase - namely, orienting the spacecraft such that the earth falls within a solid angle specified by the gimbal limits - may be discussed more quantitatively if it is assumed that the predictive capabilities of the DSIF tracking system are utilized.

Relative tangential velocities can be computed by the DSIF to an accuracy better than ±5 meters/sec.[8-1] If the error in the relative tangential velocity, V_T, is designated as ΔV_T, and assuming that $\dot{V}_T \cong 0$ between interrogations, then the resultant angular error due to incorrect prediction of the rotation of the LOS is

$$\Delta\theta_p = \frac{\Delta V_T}{R}(T_1) \quad ; \qquad (8\text{-}1)$$

where T_1 is the desired time between successive interrogations and R is the range. $\Delta\theta_p$ is given in Figure 8-6 for ranges typical of a Mars mission.

The accuracy with which the direction of the earth-to-DSV LOS vector may be determined is limited by beam steering due to atmospheric effects to about ± 1 μrad for excellent conditions and to ± 15 μrad in the presence of strong turbulence. Therefore, for T_1 less than a day or so, the DSIF may predict the direction of the earth-to-DSV LOS vector to within ± 20 μrad. Presumably the accuracy to which the lead angle can be calculated falls well within this range so that the direction of the required DSV-to-earth LOS vector can be specified to within ±20 μrad.

Finally, the errors associated with the star and sun sensors will contribute to the uncertainty in positioning the DSV transmitter something on the order of 20 to 30 μrad. The end result is that the DSV transmitter axis can be pointed to within ± 40 μrad (rms) of the correct DSV-to-earth LOS. Since the earth subtends an angle of 30 μrad at Mars ranges, the initial acquisition of the earth beacon can be programed to a large degree to speed acquisition.

8.2.6 DSV Tracking and Pointing Loops

Accurate pointing of the large massive structures such as the telescope proposed is most efficiently done by having the entire spacecraft react to the torquing of a small inertia wheel type control system. However, in certain applications, special forces may disturb the spacecraft (e. g., man motions) or it may be required to point simultaneously two telescopes in different directions. In such circumstances the telescope must be free to move relative to the spacecraft, and a means of controlling the telescope must be provided.

If the laser transmitter and receiver use common optics, a means of pointing the transmitted beam with a variable lead angle from the received beam is necessary. In general, the pointing system will be composed of the spacecraft attitude control system, a two-gimbal

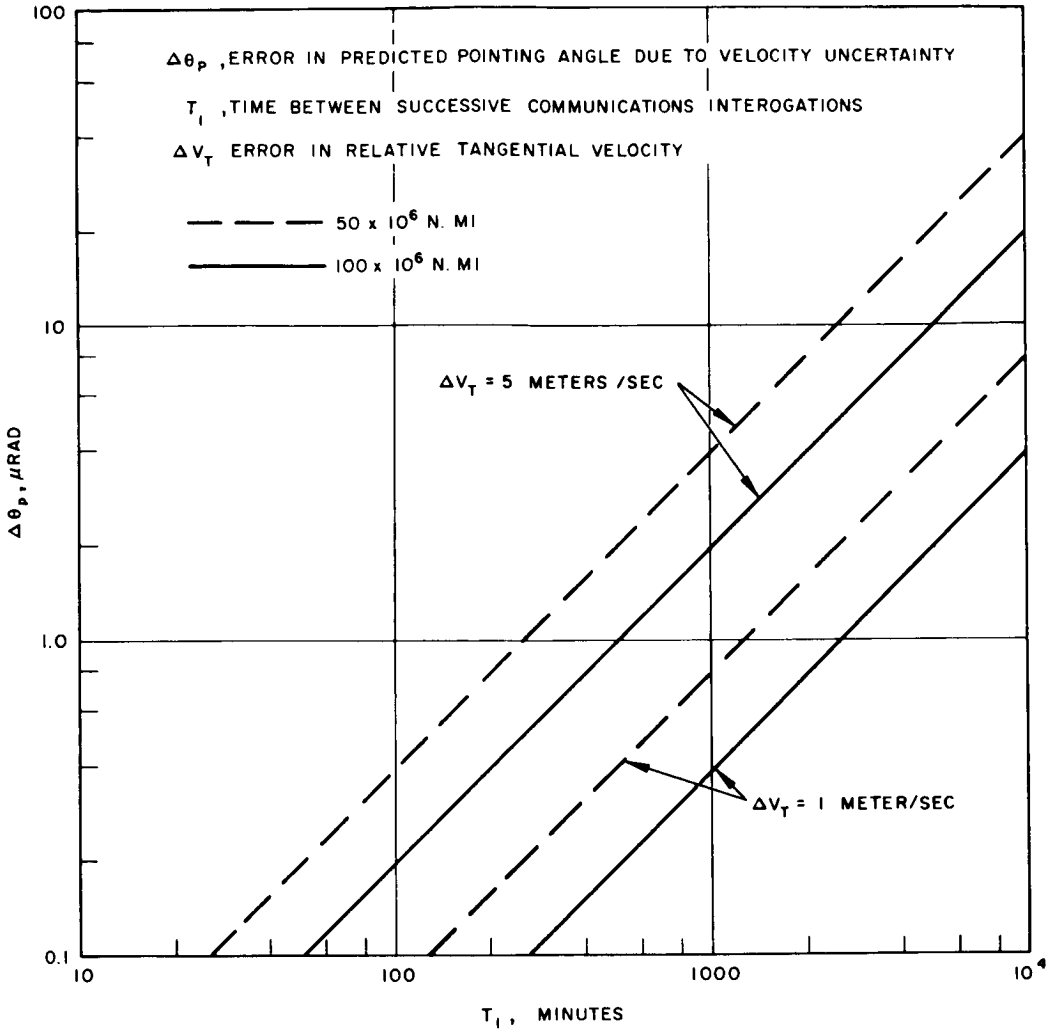


Figure 8-6. Error in predicted pointing angle time between successive versus communications interrogation.

system to orient either the entire telescope assembly or an associated mirror assembly, and an internal beam deflector within the optical assembly.

The spacecraft attitude control system serves to point the telescope in the general direction and to provide a limit cycle type of attitude control. Propellant consumption is inversely proportional to the attitude and rate limits. A very tight limit cycle requires high propellant consumption even with an inertial wheel control system.

The vehicle control system block diagrams are shown in Figures 8-4 and 8-5. A block diagram of the two gimbal system used to point the telescope is shown in Figure 8-7. Tracking error signals are used for the attitude feedback loop and tachometers are used for rate stabilization. A drawback to rate gyros is that the reference axis may be rotating. In this case the desired rate stabilization signals are proportional to the rate of deviation from the rotating reference; however, rate gyros do sense absolute angular velocities.

There are several synthetic rate schemes for stabilization which will be investigated for the next report. Three promising schemes are the phase-plane logic scheme, the differentiating gas jet scheme for vehicle attitude control, and the derived-rates scheme for the gimbal stabilization loops.[8-2]

Typical tracking rates and accelerations for a 174 day Mars mission are shown in Figure 8-8.[8-3] Due to the low tracking rates involved, a Type I servo system is considered for the telescope stabilization and tracking system.

Using the error signals derived from the optical tracking error sensors in the telescope, and information from the DSV computer and/or predictions made by the DSIF system, the beam deflecting system must give a transmitted laser beam the proper lead angle and also point the beam with an error appreciably less than the transmitted beam width. For beam widths of 1 to 50 microradians, this pointing accuracy must be within 0.1 to 5 microradians respectively. From Figure 8-8, the maximum lead angle is shown to be approximately 65 microradians. A mirror used to fold the laser beam may be gimbaled

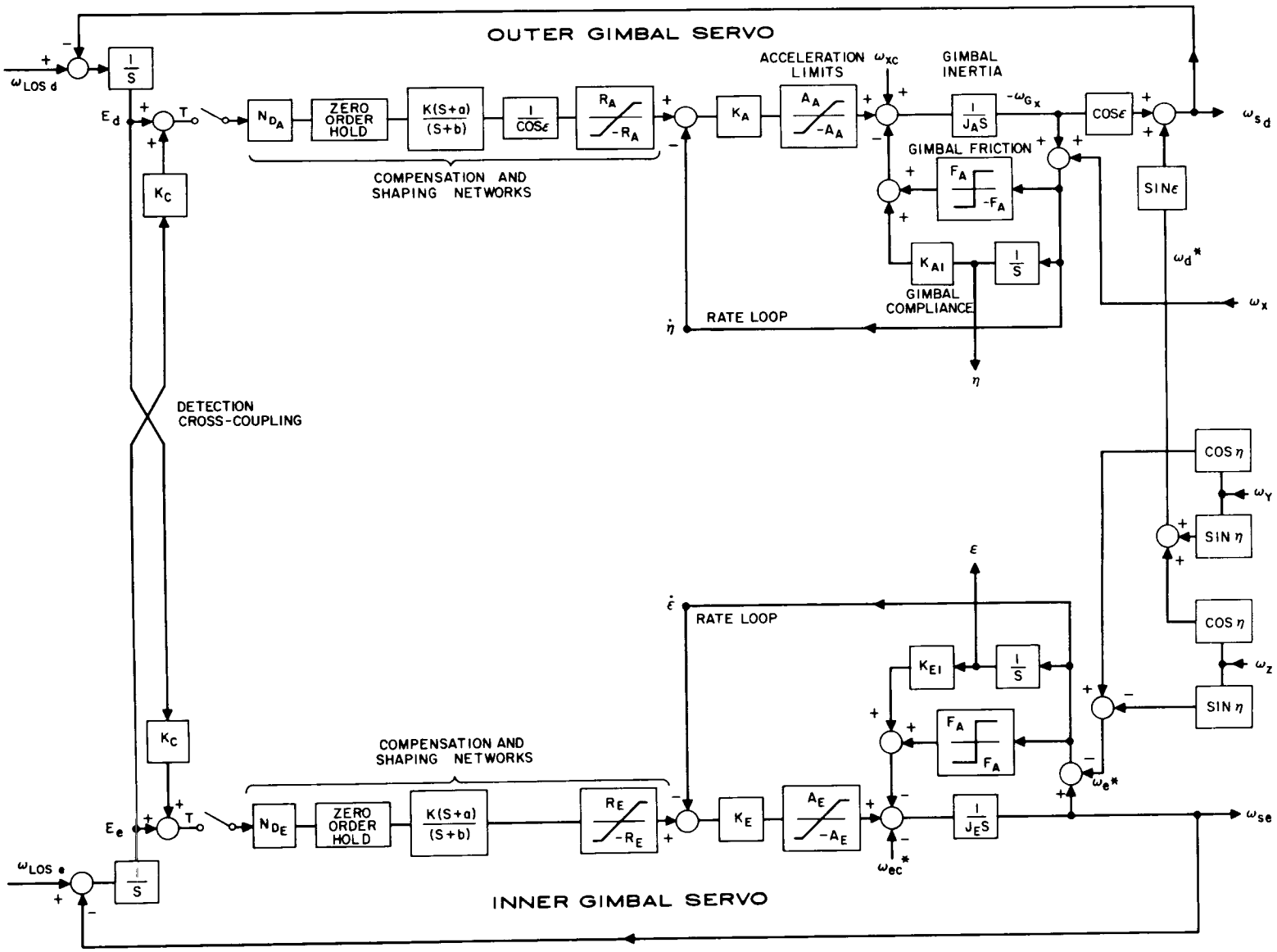


Figure 8-7. Two axis block diagram of the DSU tracking system.

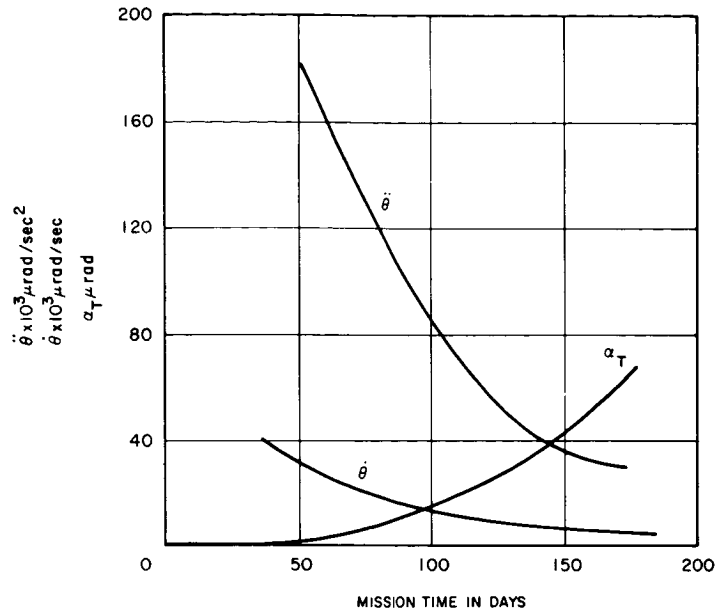


Figure 8-8. Lead angle, tracking rate, and tracking acceleration curves for a 174 day Mars trajectory.

and used for this purpose. This mirror assembly could be a two gimbal affair, or a single gimbal for the pointing angle ($\alpha_T + \epsilon_{\alpha_T}$) if the rotational component ($\beta_T + \epsilon_{\beta_T}$) is removed by a rotating prism or by attitude control of the spacecraft. It is shown elsewhere in this section that the lead angle and its perturbations are much more critical than the rotational angle. Thus, a single axis pivot to compensate for α_T might be the most logical choice.

8.2.7 Stabilization Techniques

Torquing methods for use in the communication control system are summarized in Table 8-1. Reaction jets will be required to damp large vehicle rates during initial acquisition and during periods of communication and tracking beacon loss. Electromagnetic field sensing devices are attractive; but in deep space, the fields are too weak for precise control, thus ruling out such devices. Gyroscopes are not suitable since they have extensive drift, null-offset, and reliability

Torquing Method	Major Advantages	Major Disadvantages
Gyro-torquers	Less electrical power required than with reaction flywheels	More complex than reaction flywheels; wheels run at full speed continuously; potential bearing problems
Continuous-running reaction flywheels	Feasible and simple means of momentum storage; flight demonstrated performance; precise and smooth fine pointing control	Momentum unloading required for secular disturbance torques; wheel speeds must be kept reasonable; potential bearing problems
Reaction jets	Rapid action; compressed nitrogen is attractive because it is noncontaminating and flight demonstrated; water rocket is also attractive	Stored fuel is consumed; possible leakage; many propellants can contaminate; excessive propellant usage for fine pointing
Solar dynamic pressures on movable vanes	No stored fuel expenditure	Forces too weak for precise control of synchronous attitude; complex computations required for nonsolar-oriented spacecraft
Gravitational gradient torques	No stored fuel expenditure	Fields too weak for high precision control in deep space
Controlled current coils	No stored fuel expenditure; reliable solid-state circuitry	Required field sensing; weak fields in deep space

Table 8-1. Stabilization and control system torquing techniques summary.

problems. Reaction wheels show promise for fine attitude control and will be investigated further.

Attitude sensing methods are summarized in Table 8-2. The relative merits of each sensing technique will depend on the quality of attitude control required. The earth beacon sensor is an obvious choice for this study. The stable platform is applicable to manned missions where it can be aligned and drifts corrected periodically. The fields in deep space are too weak for the ambient field sensors. Rate gyros, like platforms, when used as attitude sensing devices require drift correction and thus are not promising for unmanned deep space missions. In addition to the earth beacon sensor, the most promising instruments are the sun sensors and the star trackers. These instruments are simple, rugged, and reliable. They do however, have relatively high power and weight requirements.

The most difficult aspect of stabilization to the accuracies required for the DSV is associated with the generation of error signals of sufficient resolution and the alignment of the sensor sensitive axes with the control axes. The extreme precision control required necessitates selecting a vehicle configuration that will minimize the external disturbance torque effects, as well as minimizing internal disturbance torques caused by inertial crosscoupling, equipment motion, temperature gradients, etc. Disturbance torques due to internal moving parts can be minimized by restricting activity during the fine tracking. However, disturbance torques due to inertial crosscoupling can be significant unless care is taken to balance the vehicle such that the inertias in all three axes are approximately equal. In addition care must be taken to minimize the angular momentum stored in the vehicle.

Inertial reaction devices, which are also momentum storage devices, are attractive for steady-state control in either a fine or coarse mode and in the presence of both cyclic and secular external disturbance torques. The advantage of momentum storage is that disturbance torque effects can be transferred to the momentum storage device with minimal effects on vehicle motion. The disadvantages are; (1) the momentum must be eventually removed, and (2) the net angular

Sensing Method	Major Advantages	Major Disadvantages
Stabilized platform with gyros	Wide angle coverage; no field-of-view acquisition; rate mode for nulling tumble rates	Complex; heavy; must be updated due to drift and orbit rate; 1 to 2 year lifetime imposes questionable reliability; large
Ambient field sensors (gravity gradient, magnetic, etc.)	Simple, relatively passive sensors	Fields too weak and unpredictable in deep space, unless near planet
Sun sensors	Simple, reliable devices; no moving parts; flight demonstrated; lightweight; minimum power	Pointing accuracy presently limited to ±10 arc sec unless complicated horizon-scan attempted
Earth beacon sensor	Provides earth line-of-sight; applicable for coarse mode operation with gimballed telescope; flight demonstrated	Acquisition sequence required; weight, power requirements are large compared to sun sensors; some moving parts
Rate gyros	Lightweight; low power can be used for tumble rate nulling	When used as position signal, updating required due to drift; lifetime possibly questionable for continuous use
Star sensors	Provide accurate roll reference for coarse-fine modes; flight demonstrated	Relatively high power weight requirements; nonpassive mechanically; small acquisition field of view; potential discrimination problems

Table 8-2. Attitude sensing techniques summary

momentum vector of the device adds an additional inertial crosscoupling disturbance to the vehicle. This is in addition to that produced by unequal moments of inertia. The momentum removal requirement is easily satisfied by addition of gas jets to dump momentum; the cross-coupling disturbance can be minimized by selection of a system with a comparatively small average net angular momentum vector. For these reasons momentum storage is quite attractive, especially in the fine pointing mode.

Several types of momentum storage devices are worthy of discussion: reaction wheels (conventional or fluid), gyro-torquers, and the reaction sphere. Of the three (conventional), reaction wheels appear most attractive due to their more advanced state of development and space utilization, as well as their potential reliability advantages. According to Haeusserman [8-4] a properly designed reaction wheel stabilization system is capable of holding vehicle drift rates to within 5 μrad/min in an open loop configuration.

As will be shown subsequently, these same reaction flywheels can also be utilized during the coarse mode for control by appropriate gain and sensor switching. Several candidate motor-flywheel designs will be considered for this application. Considering reliability limitations, only ac motor structures and brushless dc motors will be considered. Included are the conventional ball-bearing supported flywheel, a gas-bearing flywheel, and a fluid flywheel system. The conventional motor flywheel appears to be the most attractive method at this time; however, a thorough investigation of the relative merits and tradeoffs of both the reaction flywheels and the gyroscopic wheels will be undertaken.

8.2.8 DSV Acquisition Subsystems

A signal flow diagram for the acquisition subsystem is shown in Figure 8-9. The detector considered is a photoelectric device which produces an electrical signal whose amplitude is proportional to the received energy. The electrical output is passed through a linear filter and demodulated. The demodulated signal is applied to a thresholding device. If the signal exceeds the threshold level it is to be concluded that the received beam is directed at the transmitter.

It is generally desirable to accomplish acquisition of the transmitter as rapidly as possible. The speed of acquisition is limited by the possibility of passing over the transmitted signal without seeing it if the solid angle is scanned too quickly, and the possibility of acquiring false signals produced by noise if the threshold is reduced to increase the sensitivity of the receiver to the transmitted signals.

A quantity $\overline{T}_a$ is defined as the mean time to acquire. The system is to be designed to minimize this number subject to various constraints imposed by other conditions. The average cost in lost time, due to false acquisitions during the interval required to scan the acquisition field once, is

$$C_o = R_f T_\ell (T_\Sigma - T_\sigma) \tag{8-2}$$

where R_f is the average rate of false target acquisitions, T_ℓ is the lost time due to acquiring a false target, T_Σ is the minimum time to scan the field of view, and T_σ is the time spent on target during a single scan. The average time needed to scan the acquisition field is then

$$\overline{T}_\Sigma = (T_\Sigma + C_o) \tag{8-3}$$

If it is determined that no signal is present when the transmitter is in the field of view, it is necessary to scan the complete acquisition field again before acquisition can be made. If the decision is "no target"

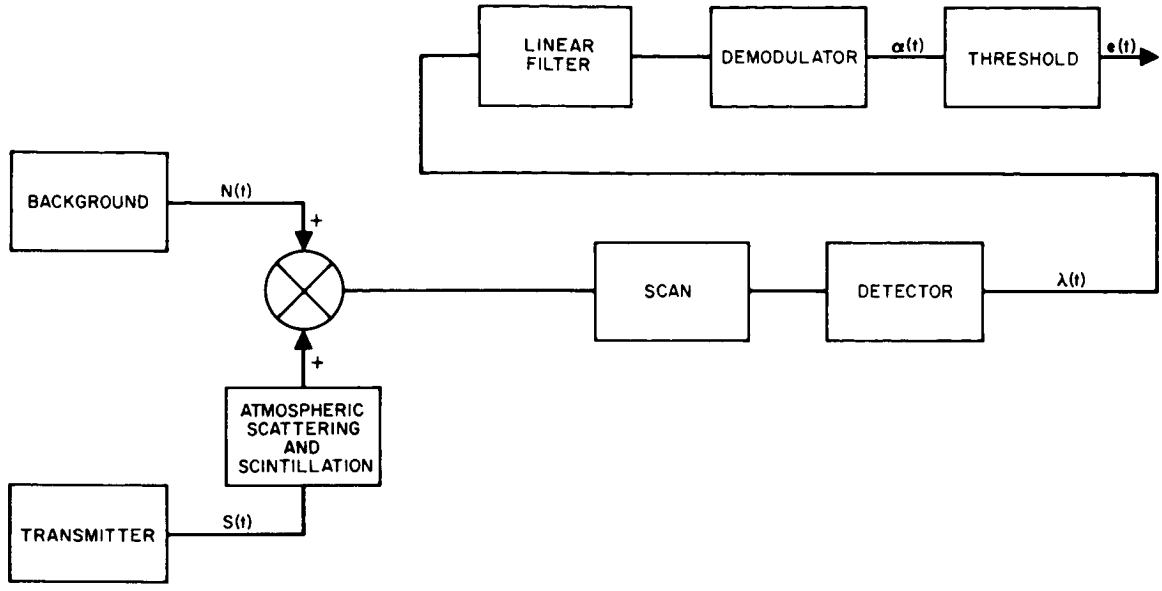


Figure 8-9. DSV Acquisition signal flow diagram.

on the second scan, the field scan is repeated, etc. If the target has equal probability P_o, of being anywhere within the field, the mean time to acquire is

$$\overline{T}_a = \frac{\overline{T}_\Sigma}{2} P_o + \frac{3}{2}\overline{T}_\Sigma P_o(1 - P_o) + \frac{5}{2}\overline{T}_\Sigma P_o(1 - P_o)^2 + \cdots$$

$$= \frac{\overline{T}_\Sigma}{2}\frac{(2 - P_o)}{P_o} \qquad (8\text{-}5)$$

The minimum time in which the field can be scanned is

$$T_\Sigma = \frac{\Sigma}{R_s}$$

where Σ is the size of the field of view and R_s the scanning rate. For a target of angular size σ,

$$T_\sigma = \frac{\sigma}{R_s} \qquad (8\text{-}6)$$

In terms of the fundamental parameters the mean time to acquire becomes

$$\overline{T}_a = \frac{\Sigma}{2R_s}\left[1 + R_f T_\ell (1 - \frac{\sigma}{\Sigma})\right]\left(\frac{2 - P_o}{P_o}\right) \qquad (8\text{-}7)$$

This assumes a uniform scan in which every element of the acquisition field is in view for the same length of time.

The quantities R_f and P_o are dependent on the noise which arrives at the thresholding device. Sources of noise are the photoelectric detector, background radiation, fluctuations in the signal and background due to atmospheric effects, and the random distribution in time of the photons which constitute the received power.

Before a particular acqusition system can be specified, the statistical properties of the output e(t) must be determined as a function of the various fixed and variable system parameters, so that R_f and P_o may be computed.

Discrimination against the background beyond that provided by an optical filter may be necessary. This can be achieved by modulating the transmitted signal at some relatively high frequency. Although some modulation is placed on the background by atmospheric effects, it decreases in amplitude as the frequency increases and should become insignificant at sufficiently high modulation frequencies.

8.2.9 DSV Tracking Loop Subsystem

A two axis block diagram of a Type I tracking control system is shown in Figure 8-7. This diagram includes sensor crosscoupling, lead restraint, coulomb friction, and sensor error signal scale factor variation in addition to acceleration (torque) and rate limiting.[8-5] It is not easy to write functional error equations for a system when it is considered realistically as a multi-axes system with crosscoupling due to the mechanical components and the inertial constraints. As can be seen from Figure 8-7, body rates enter the system via the appropriate Euler rate transformations. Not shown on the diagram are the inertial crosscoupling terms due to nonzero values of the cross products in the inertia tensor. Crosscoupling also exists as shown due to photo-detector unbalance. This crosscoupling leads to positive feedback between the two tracking loop channels.

Wire leads attached to the inner and outer gimbals cause restraining torques which produce pointing errors. These restraining torques are proportional to the respective gimbal angle displacements. Coulomb friction is a nonlinear term which is shown coupled into the torque summing junction. Because of the potential reliability, drift, and null offset problems, gyroscopes should be avoided when possible. In a two-gimbal system, the outer gimbal must be driven in such a manner as to cancel out the body rates. If stabilization loops with inertial

reference devices are not used, the required rate signals must be generated by the tracking loop. In Figure 8-7, the outer gimbal loop is shown at the top of the figure. The outer gimbal tracking loop gain, as noted, is a function of the inner loop gimbal angle. Thus, the $1/\cos\epsilon$ block must be added for compensation if the loop gain is to be kept constant. Obviously, large inner gimbal angles must be avoided. For DSV missions where the vehicle attitude control may be used to coarse point the telescope, the inner gimbal angle will remain small and the $\cos\epsilon$ may be neglected.

Tachometer feedback is included for rate stabilization and to improve the isolation from torque disturbances. The compensation block is a lag-lead network to increase the velocity constant. A disadvantage of the tachometer which is not readily apparent is the fact that body rates tend to cause the gimbal to be torqued in the direction of body motion. Such is not the case when rate gyros are used since these measure inertial rates and not relative gimbal angle rates as does a tachometer. Thus for tachometer feedback the terms ω_{xc} and ω^{*}_{ec} must be subtracted as shown. These values have to be calculated or derived by some synthetic means.

The two axis tracking system shown in Figure 8-7 is a Type I system rather than a Type II system for several reasons. Although a Type I system exhibits a steady state position error for rate command inputs while a Type II does not, this steady state position error is not always important if the transient response or settling time is superior to an equivalent Type II system, and if transient ramp disturbances are of greater importance than the steady state tracking rate requirements. For the deep space vehicle missions considered, the steady state tracking rates requirements are very small (0.03 microradians per second). Transient rate inputs, such as the vehicle limit cycle rates, will be much more demanding. These rates will occur first positively and then negatively in cyclic manner. The period between the sign reversals is a complicated function of the limit cycle deadbands, and the vehicle inertia crosscoupling terms. Thus, transient position tracking errors will determine whether or not the system is capable of

tracking and pointing to the required accuracy, and not the steady state parameters. In general, a Type I system is more stable, less sophisticated, more reliable, and has a faster response and settling time.

8.2.10 Lead Angle and Lead Angle Prediction Errors

The transmitter beam must be positioned in space without the aid of any type of feedback signal from the receiving station to null the positioning errors. The position of the transmitter pointing vector may be determined relative to the receiver pointing vector by two angles - viz., the lead angle α_T, and the rotation angle β_T measured around the receiver pointing vector (DSV to earth line-of-sight). These angles are defined geometrically in Figure 8-10. It is assumed that the lead angle α_T is small and that during the electromagnetic transit times involved, the relative velocity angular direction (aspect angle) is constant. The length of the earth-to-DSV position vector will be approximately equal to that of the DSV-to-earth position vector. If this value is R and T is the round trip transit time, then

$$T = \frac{2R}{c} \tag{8-8}$$

where c = speed of light.

The distance the earth station moves relative to the DSV in this time is

$$d_2 = V_T T \tag{8-9}$$

where V_T = the relative tangential velocity of the earth station. Note also from Figure 8-10 the angular value γ of the transmitter beam width is

$$\gamma = \frac{2d_3}{d_1} \tag{8-10}$$

with d_3 equal to the radius of the illuminated area at the receiver. For small angles, the lead angle may be written as

$$\alpha_T = \frac{d_2}{R} \tag{8-11}$$

Combining Equations 8-9 and 8-10 gives

$$R = \frac{cd_2}{2V_T} \tag{8-12}$$

Thus from Equation (8-8)

$$\alpha_T = \frac{2V_T}{c} \tag{8-13}$$

and it is seen that the lead angle is independent of the magnitude of the position vector, being only a function of the relative velocity between the transmitter and respective receiver. Clearly $V_T \ll c$ for interplanetary missions in the foreseeable future, so the small angle approximations made previously were valid assumptions.

Now the receiver on earth must be illuminated by the transmitter, thus the receiver must lie within some circle of radius d_3 determined by the power density distribution of the transmitted beam. From Figure 8-11 we see that

$$\epsilon_\beta = \frac{\left(d_3^2 - R^2\epsilon_\alpha^2\right)^{1/2}}{d_2} \tag{8-14}$$

Substitution of Equations (8-10) and (8-11) yields

$$\epsilon_\beta = \frac{1}{\alpha_T}\left[\left(\frac{\gamma}{2}\right)^2 - \left(\epsilon_\alpha\right)^2\right]^{1/2} \tag{8-15}$$

which may be written as

$$\alpha_T^2\epsilon_\beta^2 + \epsilon_\alpha^2 = \left(\frac{\gamma}{2}\right)^2 \tag{8-16}$$

Again note that the sum of the squares of the lead errors is independent of the transmission distance, but is a function of the relative velocity (due to the α_T form) and the magnitude of the transmitter beamwidth. Equation (8-16) is an ellipse as shown in Figure 8-12.

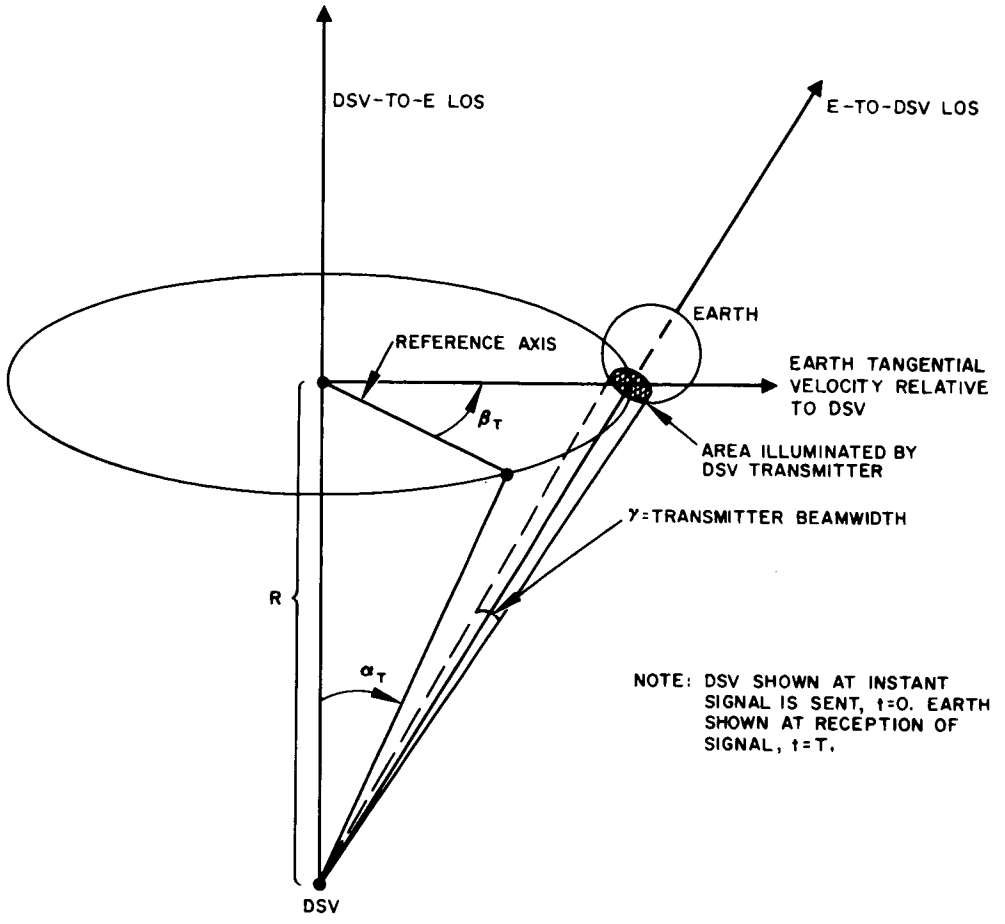


Figure 8-10. DSV transmitter lead angle geometry.

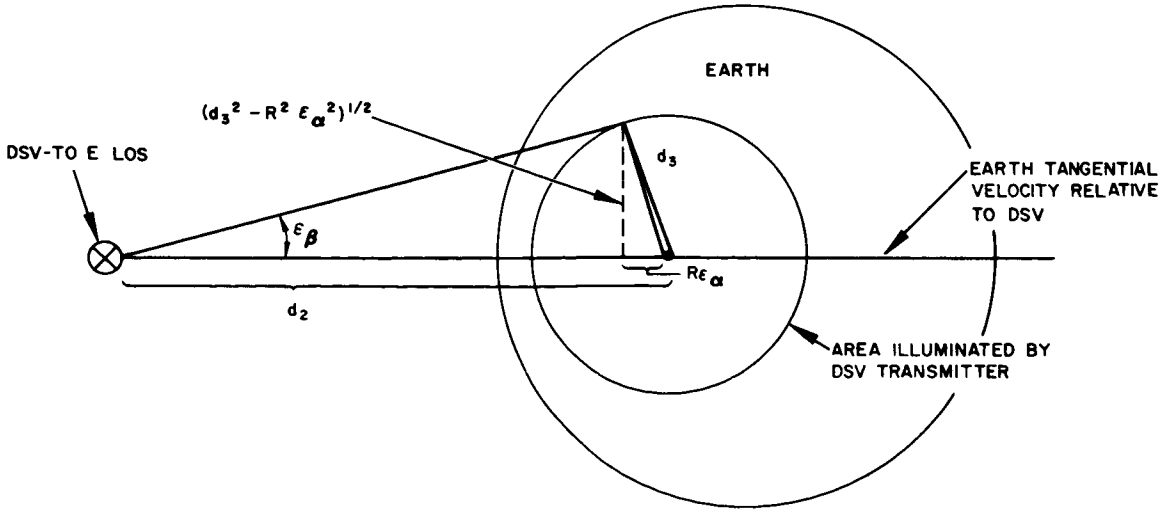


Figure 8-11. DSV transmitter lead angle error geometry.

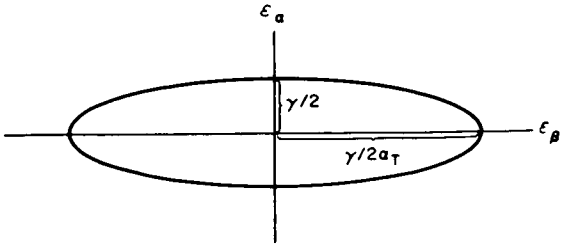


Figure 8-12. Allowable laser pointing error distribution.

Now consider the maximum tangential velocities expected for interplanetary missions, viz., 10 miles per second. From Figure 8-13, the lead angle required is approximately 100 microradians. Now from Equation (8-16)

$$\epsilon_\alpha^2 = \frac{\gamma^2}{4} - \alpha_T^2 \epsilon_\beta^2$$

Thus if

$$\frac{\gamma}{4} \gg \alpha_T \epsilon_\beta$$

then

$$\epsilon_\alpha \cong \frac{\gamma}{2}$$

For $\gamma_{min} = 1\ \mu rad$, $\alpha_{T_{max}} = 100\ \mu rad$, and $\gamma/2 = 10\alpha_T \epsilon_\beta$,

then

$$\epsilon_\beta \cong \frac{\gamma}{20\alpha_T} = \frac{1}{2} \text{ milliradian}$$

and

$$\epsilon_\alpha \cong \frac{\gamma}{2} = \frac{1}{2} \text{ microradian} .$$

Thus the allowable error on the pointing angle α_T is much less than that for tne rotation angle β_T. An intuitive way of seeing this is that the error in α_T acts on a radius equal to R while the error in β_T acts on a radius on the order of $\alpha_T R$.

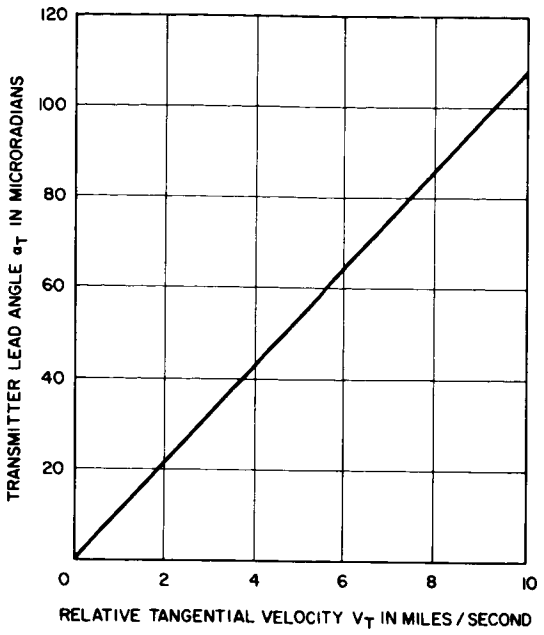


Figure 8-13. Transmitter lead angle as a function of relative lead angle.

8.2.11 Pointing Error Budgets

The composition of system pointing errors is illustrated in Figure 8-14 for a system with a tracker and transmitter mounted on the same platform with the same fine deflection tracking system. The correct pointing angle for the transmitter in inertial coordinates is given by the sum of the DSV-to-earth line-of-sight beacon angle θ_{Bd}, and the lead angle, α_T. The pointing error includes the true track error of the DSV tracker,

$$\epsilon_T = \theta_{Bd} - \theta_{td} - \psi_{fd} \qquad (8-17)$$

(where θ_{td} is the telescope axis in inertial coordinates and ψ_{fd} is the displacement of the tracker look angle from the telescope axis), the boresight error in referencing the transmitter axis to the tracker axis, and an error, $\theta_{ld} - \psi_{ld}$, in the command lead angle. In general, a classification of pointing errors into three categories is indicated:

1. Boresight and Lead Angle Errors - These errors contribute a static or nearly constant term to pointing error and are

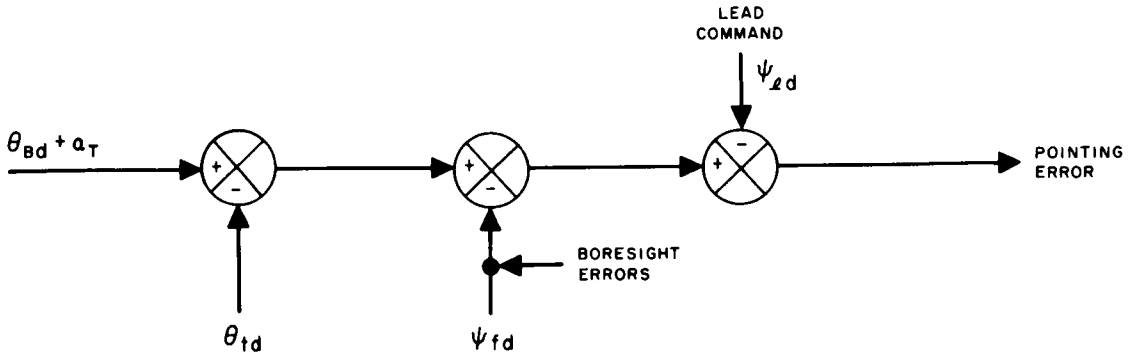


Figure 8-14. Relation of pointing error to track system quantities.

especially troublesome since they cannot be enclosed in a closed loop other than a DSV-earth closed loop. The penalty for excessive boresight errors is severe as the result is likely to be loss of transmission for an extended period of time. At ranges of 1 AU or more the round trip transit time exceeds 15 minutes and therefore correction signals from the earth will involve considerable downtime.

2. Mechanical Telescope Perturbations - The steady state response of the stabilization and tracking system may often be sufficient to suppress low frequency mechanical perturbations to a tolerable level. However, due to limited frequency response, high frequency components may cause transients in the inertial telescope pointing angle in excess of desired limits. Since these errors are sensed as apparent line of sight motions by the tracker they will be reduced by the combined action of track and stabilization loops. Thus, momentary interruption of communication may be an allowable penalty. Under maximum spacecraft maneuvers, the tracking system may not be able to contain the pointing error within prescribed limits so that a downtime penalty during extreme maneuvers may accrue.

The tracking loop need not be broken during these periods however. The system may be put in the coarse tracking mode with the telescope in a wide field of view mode. This field of view might lie somewhere between the field of view during search and the field of view during communication. Thus a tradeoff between tracking accuracy, spacecraft maneuverability, and system gain must be investigated.

3. Track Errors - The angle noise in the tracker originating from such sources as the beacon tracking sensor and the inertial reference sensor of the stabilization system may be treated by standard gaussian noise techniques. The angle noise power spectrum is modified by the closed loop response of the tracker and is included as a random term appearing in both θ_{td} and ψ_{fd}.

A position error will also be introduced due to the idiosyncracies of a Type I system. The response to a rate input of a Type I system, with unity feedback, will have a nonzero position error (final value) after the system transients have died out. This error is a function of the magnitude of the slope of the input ramp and the forward gain of the system, that is

$$\epsilon_r \cong \frac{\dot{\theta}}{|KG|} \tag{8-18}$$

For the tracking rates involved in a deep space probe the values of θ are of the order of 0.03 microradian per second. For a system gain of say 100 sec^{-1}, $\epsilon_r \cong 0.3$ nanoradian, and may be neglected.

In general these effects on the total pointing error are then written as

$$\epsilon = \epsilon_B + \epsilon_p + \epsilon_t \tag{8-19}$$

where the right-hand terms are the boresight, telescope perturbations, and tracking noise respectively. A typical system pointing accuracy specification therefore would include:

a. An absolute limit on boresight errors in the presence of thermal telescope environment, etc., and the accuracy of lead angle commands.

b. A specification on the tracking and stabilization system which reduces pointing errors as a function of deterministic mechanical disturbances of the telescope including expected rates and torques due to the attitude control system and spacecraft-earth relative motion. In final form this type of specification may include the mechanical transfer characteristics of the Deep Space Vehicle (DSV)/tracker-pointer mechanical interface.

c. A specification of the random tracking noise in the tracker to reduce loss rate in the presence of the aforementioned pointing error to the prescribed level.

Thus given the total system errors due to dynamics and boresight error, ϵ_s, the loss rate is given by

$$\Omega = 2(2\pi f_t)\left(\frac{\epsilon_o}{\epsilon_m}\right) e^{-(\epsilon_o/\epsilon_m)^2/2} \quad \text{sec}^{-1} \qquad (8\text{-}20)$$

where

f_t = closed loop bandwidth

ϵ_o = remaining system margin = beamwidth radius $-\epsilon_s$

ϵ_m = RMS tracker angle noise

Thus the tracker sensor accuracy may be specified as a function of the desired loss rate.

8.2.12 Angle Readout and Nonorthogonality Errors

In the following, the method of analysis and relations for tracking errors due to gimbal angle readout inaccuracies and misalignment of gimbal axes are presented.[8-7] Readout error is the error that exists between the actual gimbal shaft position and the indicated shaft position.

Base plate misalignment error is defined to be the angular error that arises when a gimbaled system is mounted on a reference base. The outer gimbal axis of rotation is taken as the reference for determining these mounting errors, and the mount may be misaligned with respect to a similar axis contained in the base. Since angles measured about three orthogonal axes completely specify the mounting misalignment, angles corresponding to roll ($\Delta\alpha$), pitch ($\Delta\beta$), and yaw ($\Delta\lambda$) are used for convenience.

In addition to base plate misalignment, the inner gimbal axis may not be orthogonal to the outer gimbal axis and cause pointing errors. This angle ($\Delta\gamma$) is measured in a plane orthogonal to the gimbal pointing direction.

The line-of-sight of each sensor may be misaligned with respect to the gimbal pointing direction for the mounts. These misalignment angles (Δx and Δy) are measured about an orthogonal set of axes (e and d) contained in a plane orthogonal to the gimbal pointing direction; Δx is measured about the d axis and Δy about the e axis.

These sources of error (with the exception of servo dynamic error) are due to electrical and mechanical inaccuracies (measurement and fabrication) that can only be described in a statistical manner. It is assumed that the error parameters are statistically independent and that each parameter is normally distributed with zero mean.

Since the tracking servo base motion inputs and line-of-sight tracking rates are random with time the servo errors due to these inputs can also be considered to be random variables with zero mean.

Method of Analysis. In this analysis, the distribution of the error vector resulting from the various error sources is found. The error vector is taken as the projection of the indicated pointing vector on the plane orthogonal to the ideal pointing vector.

The gimbal configuration assumed for the optical tracker consists of an outer gimbal (η) and an inner gimbal (ϵ). It is assumed that the misalignment angles are such that small angle approximations can be used (i.e., $\sin\theta = \theta$ and $\cos\theta = 1$).

Coordinate transformations can be represented pictorially thus allowing a series of coordinate transformations to be stated in concise form. Pio's method is shown in Figure 8-15 where the direction of signal flow is from left to right.

A "Pio-gram" of the coordinate transformations required to generate the error vector between the desired and actual pointing vectors for the set of error angles is given in Figure 8-15 where

η = desired outer gimbal angle

$\Delta\eta$ = outer gimbal deviation angle

ϵ = desired inner gimbal angle

$\Delta\epsilon$ = inner gimbal deviation angle

$\Delta\alpha$ = base misalignment roll angle

$\Delta\beta$ = base misalignment pitch angle

$\Delta\lambda$ = base misalignment yaw angle

$\Delta\gamma$ = gimbal nonorthogonality angle

Δx = sensor misalignment angle (about d axis)

Δy = sensor misalignment angle (about e axis)

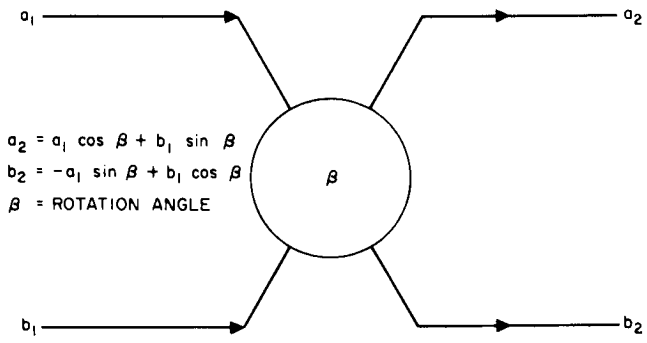


Figure 8-15. Pio's method of pictorial representation of coordinate transformation.

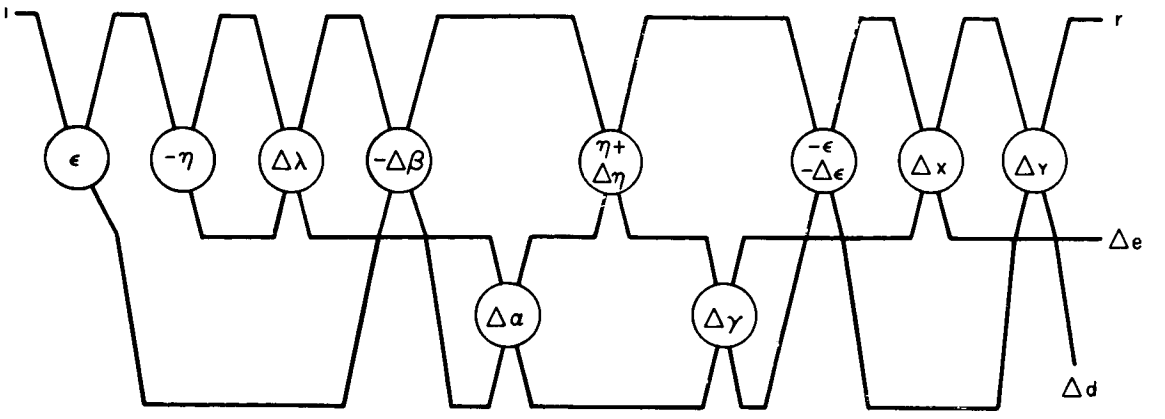


Figure 8-16. Error vector "Pio-gram".

From Figure 8-16 it is determined that

$$\Delta e = \Delta\eta \cos\epsilon - \Delta\alpha \sin\epsilon \cos\eta - \Delta\beta \sin\epsilon \sin\eta$$
$$- \Delta\lambda \cos\epsilon - \Delta\gamma \sin\epsilon - \Delta x \qquad (8\text{-}21)$$

$$\Delta d = \Delta\epsilon - \Delta\alpha \sin\eta + \Delta\beta \cos\eta + \Delta y$$

If the sensor is aligned (i.e., no servo errors) with both gimbal angles at zero, Equations (8-21) reduce to

$$-\Delta\gamma - \Delta x = \Delta_{ae}$$
$$\Delta\beta + \Delta y = \Delta_{ad} \qquad (8\text{-}22)$$

where

Δ_{ae} = outer gimbal axis alignment error

Δ_{ad} = inner gimbal axis alignment error

Substituting Equations (8-22) into Equations (8-21) yields

$$\Delta e = -\Delta\eta \cos\epsilon - \Delta\alpha \sin\epsilon \cos\eta - \Delta\beta \sin\epsilon \sin\eta + \Delta\lambda(1 - \cos\epsilon) - \Delta\gamma \sin\epsilon + \Delta_{ae} \tag{8-23}$$

$$\Delta d = \Delta\epsilon - \Delta\alpha \sin\eta - \Delta\beta(1 - \cos\eta) + \Delta_{ad}$$

Using one sigma values for the deviation angles, the 1σ values for the error vectors can be computed by the propagation of variances. The resulting expressions are:

$$\Delta e_{1\sigma} = \left[\Delta\eta_{1\sigma}^2 \cos^2\epsilon + \Delta\alpha_{1\sigma}^2 \sin^2\epsilon \cos^2\eta + \Delta\beta_{1\sigma}^2 \sin^2\epsilon \sin^2\eta + \Delta\lambda_{1\sigma}^2 (1 - \cos\epsilon)^2 + \Delta\gamma_{1\sigma}^2 \sin^2\epsilon + \Delta_{ae_{1\sigma}}^2\right]^{1/2} \tag{8-24}$$

$$\Delta d_{1\sigma} = \left[\Delta\epsilon_{1\sigma}^2 + \Delta\alpha_{1\sigma}^2 \sin^2\eta + \Delta\beta_{1\sigma}^2 (1 - \cos\eta)^2 + \Delta_{ad_{1\sigma}}^2\right]^{1/2}$$

and

$$\rho = \frac{\Delta\alpha_{1\sigma}^2 \sin\epsilon \sin\eta \cos\eta + \Delta\beta_{1\sigma}^2 \sin\epsilon \sin\eta (1 - \cos\eta)}{\Delta e_{1\sigma} \Delta d_{1\sigma}} \tag{8-25}$$

where ρ is the correlation coefficient. If the base misalignment is numerically equal to $\Delta\alpha_{1\sigma}$, and the alignment error 1σ values are the same for both axes ($\Delta a_{1\sigma}$), Equations (8-24) reduce to

$$\Delta e_{1\sigma} = \left[\Delta\eta_{1\sigma}^2 \cos^2\epsilon + \Delta\alpha_{1\sigma}^2(2 - 2\cos\epsilon) + \Delta\gamma_{1\sigma}^2 \sin^2\epsilon + \Delta\alpha_{1\sigma}^2\right]^{1/2}$$

(8-26)

$$\Delta d_{1\sigma} = \left[\Delta\eta_{1\sigma}^2 + \Delta\alpha_{1\sigma}^2(2 - 2\cos\eta) + \Delta a_{1\sigma}^2\right]^{1/2}$$

and the correlation coefficient to

$$\rho = \frac{\Delta\alpha_{1\sigma}^2 \sin\epsilon \sin\eta}{\Delta e_{1\sigma} \Delta d_{1\sigma}} \quad . \tag{8-27}$$

For the case where the optics are not gimbaled but fixed to the spacecraft, and where a gimbaled mirror is used for coarse pointing, Equations (8-24) through (8-27) must be modified accordingly. Also the optical system which is used to provide the transmitter lead angle will contribute to the error. General error equations cannot be written for the optical system until a decision has been made as to what type of gimbal configuration will be used and where the telescope line-of-sight will be interrupted for folding. Further, the lead angle pointing mirror will contribute an error dependent on the size and type of collimation optics used. The angle of a lead angle mirror will be

$$\alpha_m = \frac{d_2}{2d_1}\alpha_T \quad . \tag{8-28}$$

Thus, the pointing error due to a lead angle mirror adjustment is

$$\epsilon_{\alpha_T} = \frac{2d_1}{d_2}\epsilon_{\alpha_m} \quad . \tag{8-29}$$

The effect of the mirror itself is to double the pointing error. The effect of the ratio of the beam expansion optics diameter to the diameter of the recollimation optics, d_1/d_2, is to decrease the pointing error.

If this ratio is 40:1, as might be required for beam widths of several microradians, then the error tolerance on the mirror is increased by a factor of twenty over the pointing error tolerance.

The lead angle pointing mirror error, in conjunction with some type of gimbaled optical system, will be discussed in the next report. Also to be considered will be the vehicle attitude control. The gimbal angle magnitudes, which depend on the vehicle attitude and rate limit cycle values, will also be derived and suitable system restraints developed.

8.2.13 Sensor Boresight Errors

In a typical image sensor (Vidicon, Orthicon, image dissector) the nonlinearities relating beam position to sweep voltagle produce a tracking error which is proportional to the angular track field. In most cases this error is of the order of 1/2 to 1 percent. The quadrant photodetector depends for angle accuracy on the balance of energy in the blurred image of a point target among the four sections of the field of view. Even in the absence of noise, the null accuracy is limited by the degree to which the gains of the four channels can be balanced. Figure 8-17 shows the boresight error as a function of gain imbalance in a single channel. Thus in order to reduce boresight errors to less than 1 percent of the linear response portion of the field of view, gain imbalance must be held to less than 10 percent over the dynamic range of signal levels.

8.2.14 Earth Tracking Subsystem

The tracking subsystem for an earth based acquisition and tracking system is considerably simplified by the extremely low angular target rates. The diurnal rate (15 arcsec/sec) may be accurately removed by a clock driven mechanism as in astronomical telescopes. The track loop will thus have a low time constant and typically provide integration of the input angle error for 1 sec or more without appreciable lag angles.

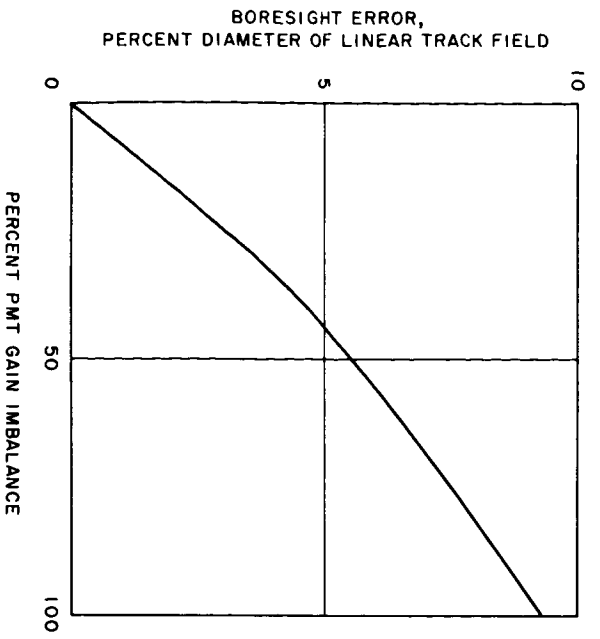


Figure 8-17. Quadrant detector boresight errors.

The track accuracy of the ground based sensor is determined jointly by the sensor noise and the angle noise spectrum of the scintillation phenomena.

The total rms angle noise is then given in terms of the track sensor and scintillation noise by

$$\epsilon_N^2 = \int_0^\infty df[W\eta_t(f) + W\eta_s(f)] \frac{|Gc(f)|^2}{|1+Gc(t)|} \tag{8-30}$$

where Gc(f) is the open loop frequency transfer function. For a typical Type I track loop and white sensor noise

$$\int_0^\infty df\, W\eta_t(f) \frac{|Gc(f)|^2}{|1+Gc(f)|^2} = \frac{K_V}{4} W\eta_t(0) \tag{8-31}$$

where K_V is the velocity constant.

The choice of velocity constant can be varied during the mission to improve performance.

For a Mars mission the angular rates at ranges greater than 10^7 n.mi. are less than 1×10^{-2} arcsec/sec so that several seconds of smoothing will not cause any appreciable servo lags. Thus typical values of K_V will be on the order of 0.05 sec^{-1} for 0.2 arcsec steady state position error.

8.2.15 Pointing Error Budget for an Earth Station

The major sources of pointing error in the ground tracking and transmitting system are:

1. Atmospheric Scintillation - Angle scintillation due to random phase errors introduced into the spacecraft beacon by the atmosphere.
2. Boresight and Lead Angle Errors - The pointing errors due to misalignment of the earth beacon and earth tracker and errors in introducing the beacon lead angle are of the form of bias errors
3. Tracker Errors - Errors introduced in the tracker system such as tracker noise and track sensor boresight errors.

The total pointing error is thus

$$\epsilon = \epsilon_s + \epsilon_B + \epsilon_t \tag{8-32}$$

The track and scintillation errors define random angle noise whose effect may be computed from a knowledge of the angle noise power density and the track loop characteristics.

Denoting by ϵ_N the rms sum of the scintillation and tracker noise contributions to the tracking error the rate at which an earth beacon of angular radius ϵ_o is displaced from the spacecraft is

$$\Omega \simeq 2\pi f_t x e^{-x^2/2} \text{ per sec} \tag{8-33}$$

where

$x = (\epsilon_o - \epsilon_B)/\epsilon_N$, and f_t is the closed loop bandwidth of the earth tracker.

8.2.16 Earth Satellite Considerations

The degradation in system performance due to the effects of the atmosphere may be severe. One obvious solution is to orbit a relay satellite thus eliminating all atmospheric disturbances save those encountered in the communications link between the earth and the satellite. Such a "solution", of course introduces other problems which must be traded off against the atmospheric effects.

The selection of possible orbits for the near earth relay satellites will be influenced by lifetime and thermal variation control as well as minimizing the number of handovers required to support various deep space missions. The most appropriate configuration for the near earth relay station is a body mounted telescope with an inertially stabilized attitude control system for the satellite. One possible configuration is that of the OAO with reaction wheel stabilization. Since the near earth relay would be required only for support of deep space missions, the extremely low inertial angular rates at ranges on the order of 1AU will allow the very long smoothing times used in this type of system to achieve high tracking accuracy. By allowing a fine deflection mechanism with several arc seconds dynamic range, the period between attitude corrections in the track mode may be extended to several days and probably will not be limited by tracker dynamic considerations.

The major interfaces are with the DSIF which must supply initial acquisition and lead angle information to the acquisition and tracking system. This requires that the inertial attitude of the spacecraft be telemetered to the DSIF with sufficient accuracy to determine the lead angle corrections within the allowable tolerances. This requires attitude readouts to somewhat better than one minute of arc.

The pointing error budget is established as the sum,

$$\epsilon = \epsilon_B + \epsilon_p + \epsilon_t \qquad (8\text{-}34)$$

of boresight, spacecraft perturbations, and track errors. The tolerable level of each is determined from the loss rate equation.

8.3 STAR TRACKERS AND SENSORS

8.3.1 General

The establishment of a communications link begins with the acquisition and tracking of the sun and a given star. Having thus defined a set of orthogonal axes in inertial space, the telescope is pointed toward the earth and begins searching the associated error volume. A brief survey of the representative state-of-the-art of star and sun sensors has been undertaken and will be extended in the next quarter. The accuracies quoted in this section are representative but are not the ultimate for the present state-of-the-art, for it should be possible to develop star trackers capable of defining the line of sight to within the predictive capabilities of the DSIF.

The first problem in the design of an optical tracker is to code the position of the target signal in the field of view (FOV). Such tracking systems usually generate tracker information by a scanning process that produces relative motion between the optically-formed image and a reticle located in the image plane. Of the three coding methods generally used - viz. AM, FM and PCM (pulse code modulation) - only the latter two are applicable here, since AM has an inherent central dead zone which severely limits the system's accuracy. The FM and PCM methods are comparable in size, complexity, and performance, but further consideration must be given to their relative background discrimination capabilities.

Discrimination is usually accomplished in one or a combination of three ways: (a) spatially, (b) spectrally, and (c) temporally. A combination of spatial and spectral discrimination is appropriate here.

Table 8-3 lists six representative instruments and their characteristics.

8.3.2 Sun Sensors

Construction of sun sensors with an accuracy of better than 10 arc seconds is not now within the state-of-the-art. The possibility exists of going to a more complicated system such as a limb tracker,

Instrument	Performance Data	Burden Parameters
ITT Federal Labs OAO Boresighted Star Tracker	1. Star Magnitude Sensitivity +6 2. FOV 10 Arc Min 3. RMS Accuracy 1.5 Arc Sec 4. Gimballed (Electronically) ±1.5 deg 5. Two-Axis Capability	1. Weight 25 lbs 2. Power Requirements ±28V dc 7.7 watts 3. Size 3" (dia) × 15" Electronics: 5" × 11" × 12"
ITT Federal Labs Dual Mode Star Tracker	1. Star Magnitude Sensitivity +3 2. FOV $8^{o} \times 8^{o}$ Acquisit. 32' × 32' Track 3. Accuracy (RMS) ±5 Arc Sec 4. Not Gimballed 5. Two-Axis Capability	1. Weight 9.5 lbs 2. Power 8 watts (26V dc) 3. Size 5" × 10.5" × 5"
ITT Federal Labs Electro-Optical Sensing Head for Star Tracking	1. Sensitivity +2.5 2. FOV $1^{o} \times 1^{o}$ 3. Accuracy (RMS) ±9 Arc Sec 4. Not Gimballed 5. Two-Axis Capability	

Table 8-3. Star trackers.

Instrument	Performance Data	Burden Parameters
Kollsman Inst. Co. Gimballed Seeker	1. Sensitivity +2 2. FOV $1^{o} \times 1^{o}$ 3. RMS Accuracy ±5 Arc Sec 4. Gimballed (Mechanical) $\pm 4.3^{o}$ 5. Two-Axis Capability	1. Weight 42.5 lbs 2. Power 29 watts 3. Size N. A.
Bendix Corp. Star Sensor	1. Sensitivity >+2.1 2. FOV $10^{o} \times 10^{o}$ 3. Accuracy $<0.1^{o}$ 4. Not Gimballed 5. Single Axis	1. Weight 5.0 lbs 2. Power 8 watts 3. Size 4" × 7" × 10"
ITT Federal Labs Sensor Design Proposed for SERT-III	1. Sensitivity >2.1 2. FOV $10^{o} \times 10^{o}$ 3. Accuracy $<0.012^{o}$ 4. Single Axis	1. Weight 8.0 lbs 2. Size 2 1/4" (dia) × 8.5"

Table 8-3 (continued). Star trackers.

but the attendant difficulities and development problems seem to outweigh the advantages of having a more accurate sensor.

8.4 ACQUISITION AND TRACKING/BURDEN RELATIONSHIPS

To a first approximation the intrinsic cost of the optical tracking system is related to the diameter of the entrance aperture. Figure 8-18 plots the weight of typical electro-optical tracking instruments as a function of aperture size. These weights include both the gimbals, mechanically scanned equipment, and the electro-optical detectors. If extremely large aperture lightweight optics are body mounted to a platform such as a satellite whose whole orientation may be changed, the weight of the basic telescope can be very much less than indicated. The second curve on Figure 8-18 indicates this trend of weight for a body fixed telescope exclusive of gimbals, servo, and pointing weights. The expense of constructing such telescopes varies greatly depending upon the particular requirements of the apparatus. As a crude approximation one may estimate a cost of the order of $500 per pound for any of these telescopes.

It is somewhat early in the study to be more exacting on the system burden parameters. It has been shown in this report that several systems are not feasible for a deep space communication system while others seem to be logical choices. Until such parameters as types of beacon modulation, general mission configurations, basic optical designs, sensor configurations and atmospheric models have been investigated more fully, it will not be possible to define accurately the system burden relationships. It is anticipated, however, that the next report will contain some basic trade-off parameters for the number and types of detectors, optical gimbals (types and sizes), synthetic rate devices, and momentum storage devices.

8.5 NOMENCLATURE

DSV	deep space vehicle
LOS	line-of-sight
A_A, A_E	gimbal acceleration limits

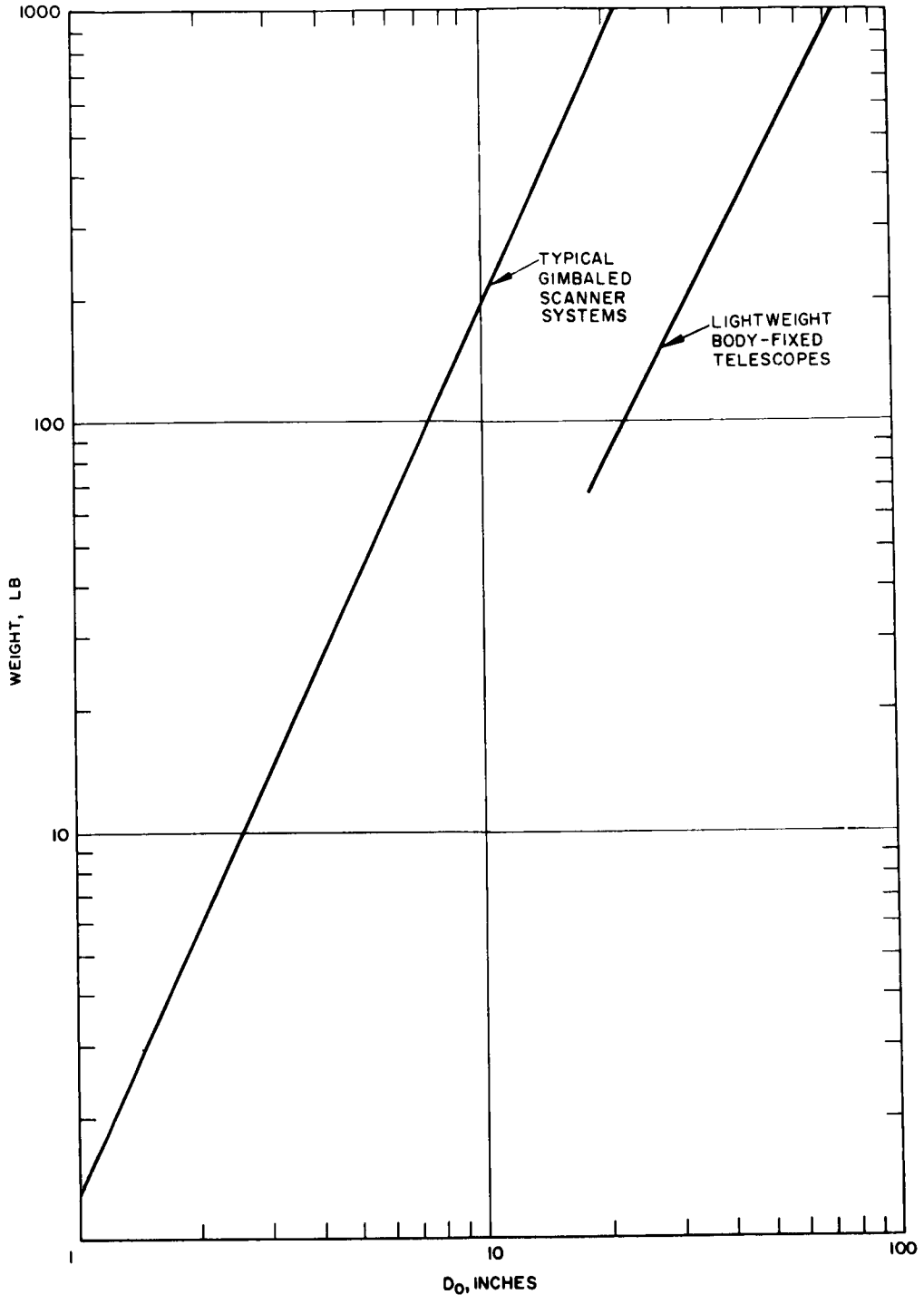


Figure 8-18. Assumed weight characteristic of different sized electro-optical apparatus.

A_D	area of detector
C	velocity of light
C_o	average cost of false acquisitions in interval
d	reference axis in the detector plane
d_1	DSV laser beam diameter
d_2	DSV telescope aperture
e	reference axis in the detector plane
e(t)	detector output
E	ratio of the effective scattering cross section to the geometric cross section of the scattering medium
f_t	closed loop bandwidth
F_A	static friction
G(f)	open loop frequency transfer function
H_m	effective height of the atmosphere for molecular scattering
H_p	effective height of the atmosphere for large particle scattering
I(o)	radiant intensity initially
I(s)	radiant intensity at point s
J_A	inner gimbal inertia
J_E	outer gimbal inertia
K_l, K_A, K_E	gain constants
K_{Al}, K_{El}	compliance
K_c	detector cross-coupling constant
K_v	velocity constant
N_a	density of scattering coefficient
N_{DA}	detector gain
P_o	probability of detecting the transmitter when it is in the field of view
P_R	received power
D_T	transmitted power
r	radius of scattering particles
R	slant range
R_E	radius of the earth

R_f	average rate of false target acquisition
R_s	minimum time to scan acquisition field
T	round trip transit time
$\overline{T}_a$	mean time to acquire
T_e	lost time due to a false acquisition
T_ϵ	scan rate
T_σ	time on target per scan
$\overline{T}_\epsilon$	average time required to scan acquisition field once
V	horizontal visibility in the atmosphere
V_T	DSV-Earth relative tangential velocity
Δx	sensor misalignment angle (about d axis)
Δy	sensor misalignment angle (about e axis)
$\Delta\alpha$	base misalignment roll angle
α_m	DSV laser folding mirror angle
α_T	trunion lead angle
$\Delta\beta$	base misalignment pitch angle
γ	transmitter beam width
$\Delta\gamma$	gimbal nonorthogonality angle
ϵ	inner gimbal Euler angle
$\Delta\epsilon$	inner gimbal deviation angle
ϵ_B	boresight error
ϵ_N^2	total rms angle noise
ϵ_p	telescope perturbation error
ϵ_r	rate tracking error
ϵ_t	tracking noise error
$\epsilon_{\alpha T}$	laser trunion pointing error
$\epsilon_{\beta T}$	laser shaft pointing error
η	outer gimbal Euler angle
$\Delta\eta$	outer gimbal deviation angle
θ	beam width angle
θ_{td}	angle of telescope axis from inertial reference
θ_z	zenith angle
λ	wavelength of radiation

$\Delta\lambda$	base misalignment yaw angle
μ	index of refraction of the air
σ	field of view during acquisition (Steradians)
Σ	size of acquisition field (Steradians)
$\sigma(s)$	atmosphere attenuation coefficient
$\sigma_{m(s)}$	molecular scattering coefficient
$\sigma_{p(s)}$	particle scattering coefficient
ψ_{fd}	displacement angle of the tracker pointing vector from the telescope axis
$\omega_x, \omega_y, \omega_z$	spacecraft body rates (inertial)

8.6 REFERENCES

8.1 Calculated from information contained in JPL Tech. Mem. No. 33-83, April 24, 1964, and Aeronautic and Astronautics, June, 1965, page 32.

8.2 Leondes, C. T., et al, Guidance and Control of Aerospace Vehicles, McGraw-Hill Book Comapny, Inc, New York, 1963, p. 415.

8.3 Proposal for a Preliminary Design Study for an Optical Technology Appollo Extension System, Hughes Aircraft Company, Technical Proposal SSD5324P, Vol. I, June 1965.

8.4 Haeusserman, ARS Journal, Feb., 1962, page 189.

8.5 Rue, A. K., "LORS Track Loop Simulation Studies," Hughes Aircraft Co., Dept. Memo No. 2783.2/60, 3 November 1965.

8.6 Rue, A.K., Fish, J., "Confidence Limits for the Pointing Error of Gimballed Sensors," Hughes Aircraft Company, Dept. Memo No. 2783.2/57, 29 September 1965.

8.7 Pio, R. L., "Symbolic Representation of Coordinate Transformation," Western Electronics Convention Paper, 1963.

12.0 BACKGROUND RADIATION AND ATMOSPHERIC PROPAGATION

12.1 INTRODUCTION

The effect of external noise is of great significance in determining ultimate communication system performance. Such information as is currently available concerning external noise sources relevant to optical and microwave space communication systems is summarized in this section. The principal types of noise encountered are:

1. Cosmic noise, originating outside the solar system
2. Terrestrial noise, originating from the earth or its surrounding atmosphere
3. Solar system noise, other than terrestrial, originating from the sun, the planets, or their satellites.

Unless otherwise stated, all measurements of extraterrestrial radio and optical background are corrected for atmospheric attenuation and refraction.

Propagation of radio frequency and optical radiation through the terrestrial atmosphere including the effects of atmospheric anomalies in introducing phase and polarization changes, beam scanning, and scintillation is considered and its implication with regard to system performance is discussed.

12.2 RADIO FREQUENCY BACKGROUND

12.2.1 Comparison of Radio Frequency Noise Sources[1]

Antenna Noise Power and Noise Temperature:

Relative contributions to microwave receiving system noise may be conveniently compared in terms of noise temperature. The concept of noise temperature is related to the Rayleigh-Jeans approximation for power density radiated by a black body cavity. The power density emanating from a unit solid angle per cycle of bandwidth (watts $m^{-2} ster^{-1} (cps)^{-1}$)

is called brightness. The brightness is given rigorously from Planck's equation by

$$B' = \frac{2hf}{\lambda^2 e^{\frac{hf}{kT}} - 1} \qquad (12\text{-}1)$$

f = frequency (cps)

h = Planck's constant (6.624×10^{-34} joule sec)

k = Boltzmann's constant (1.38×10^{-23} joules/^{o}K)

T = temperature (^{o}K)

λ = wavelength (m)

The Rayleigh-Jeans approximation to Planck's equation gives

$$B = \frac{2kT}{\lambda^2} \qquad (12\text{-}2)$$

which is valid for $hf/kT \ll 1$.

Consider the calculation of the noise power available to a receiving antenna in the frequency interval f to $f + \beta$, where β is the bandwidth (cps). The contribution to the available noise power from an elemental solid angle $d\Omega$ is obtained by multiplying the effective area of the antenna in the direction of the solid angle, by the incident power density (watts m^{-2}) having the same polarization as the antenna and originating from that elemental solid angle.

That is

$$dP = \{A(\theta_1, \phi_1)\}\{\gamma\beta B(\theta_1, \phi_1)\, d\Omega\} \qquad (12\text{-}3)$$

where

dP = the elemental available power received by the antenna from the solid angle d (watts)

= the bandwidth in question (cps)

$A(\ _1, \ _1)$ = the effective area of the antenna in the direction of the solid angle d (m^2)

$B(\theta_1, \phi_1)$ = the brightness in the direction of the solid angle $d\Omega$. (watts m^{-2} $ster^{-1}$ cps^{-1})

γ = the polarization coefficient ($\gamma = 1/2$ if the receiving antenna is linearly polarized and the incident electromagnetic radiation is randomly polarized, since only 1/2 the power in a randomly polarized electromagnetic wave can be intercepted by a linearly polarized antenna).

θ_1 and ϕ_1 = are particular values of the spatial coordinate angles θ and ϕ (i.e., the values of θ and ϕ in the direction of the solid angle $d\Omega$).

Actually Equation (12-3) can be considered to define the effective area of an antenna.

To obtain the total available power P, the contribution from all directions must be added. This superposition of power is permissible provided that the contributions from the various directions are uncorrelated as is usually the case for noise. Thus

$$P = \int\int_{4\pi} \gamma\beta A(\theta, \phi) B(\theta, \phi)\, d\Omega \qquad (12\text{-}4)$$

where

$A(\theta, \phi)$ is the effective area of the antenna as a function of the direction and will be referred to as the effective area function of the antenna

and

$B(\theta, \phi)$ is the brightness of the antenna surroundings as a function of direction and will be referred to as the brightness function of the source.

The antenna gain function is defined as

$$G(\theta, \phi) = \frac{P(\theta, \phi)}{P_T \big| 4\pi} = \frac{4\pi P(\theta, \phi)}{\int\int_{4\pi} P(\theta, \phi)\, d\Omega} \qquad (12\text{-}5)$$

where

$P(\theta, \phi)$ = the power radiated per unit solid angle as a function of the coordinates

P_T = the total power radiated.

The effective area function of an antenna is related to the antenna gain function $G(\theta, \phi)$ by

$$A(\theta, \phi) = \frac{\lambda^2}{4\pi} G(\theta, \phi) \tag{12-6}$$

In the direction of maximum gain (12-6) becomes

$$A = \frac{\lambda^2}{4\pi} G \tag{12-7}$$

where

G = the maximum antenna gain

A = the maximum effective area.

The antenna beamwidth Ω_a is defined as

$$\Omega_a = \frac{4\pi}{G} \tag{12-8}$$

The beamwidth Ω_a is the solid angle through which all the power radiated from an antenna would stream, if the power per unit solid angle equalled the maximum value over the beamwidth. Inherent in the use of beamwidth is the assumption that the gain and effective are functions are constant over the beamwidth at G and A respectively and zero elsewhere. Furthermore, the concept is applicable only to antennas that have only one principle lobe in the radiation pattern, a case of considerable importance.

The derived expression for the available power is general. If the antenna is placed in a black body cavity the brightness function is constant and is given by (12-2) (classically) or (12-1)(quantum mechanically). Using the classical equation and $\gamma = 1/2$, the available power is

$$P_b = \int\int_{4\pi} \frac{1}{2} \beta A(\theta, \phi) \frac{2kT}{\lambda^2} \, d\Omega \tag{12-9}$$

If the effective area function is expressed in terms of the gain function this reduces to

$$P_b = \frac{kT\beta}{4\pi} \int\int_{4\pi} G(\theta, \phi)\, d\Omega \qquad (12\text{-}10)$$

Since gain function of an antenna integrated over 4π steradians is 4π, the available power in a black body cavity is

$$P_b = kT\beta \text{ (classically)} \qquad (12\text{-}11)$$

and

$$P_b' = \frac{hf\beta}{e^{\frac{hf}{kT}} - 1} \text{ (quantum mechanically).} \qquad (12\text{-}12)$$

Consider the problem of calculating the noise power available to an antenna in its actual surroundings. For the lower frequencies, the classical approximation for the brightness in a black body cavity is sufficiently accurate for almost all practical purposes. Furthermore, if the actual environment exhibits a variation of brightness and if the antenna exchanges radiation on a black body basis (i.e., none of the radiation leaving the antenna ever returns to it) then Equation (12-10) is still valid provided that the directional variation of the brightness is ascribed to the temperature. Hence, provided the classical approximation is valid and assuming γ is 1/2, the power available to an antenna in its actual environment is

$$P = k \frac{1}{4\pi} \int\int_{4\pi} G(\theta, \phi) T(\theta, \phi)\, d\Omega \qquad (12\text{-}13)$$

The quantity $T(\theta, \phi)$ is the brightness temperature as a function of the coordinates. In practice, the antenna and its surroundings do not necessarily radiate like black bodies and hence, $T(\theta, \phi)$ is replaced by an effective brightness temperature $T_e(\theta, \phi)$. The effective temperature takes the emissivity and reflectivity of the antenna environment into

consideration. Replacing $T(\theta, \phi)$ in (12-13) by $T_e(\theta, \phi)$ we have for the available power

$$P = k\beta \frac{1}{4\pi} \iint_{4\pi} G(\theta, \phi) T_e(\theta, \phi)\, d\Omega \qquad (12\text{-}14)$$

Comparing this equation with (12-10) we can define an effective antenna noise temperature T_{eff} as

$$T_{eff} = \frac{1}{4\pi} \iint_{4\pi} G(\theta, \phi) T_e(\theta, \phi)\, d\Omega \qquad (12\text{-}15)$$

It is apparent that if the antenna were immersed in a black body cavity whose temperature is T_{eff}, then the noise power which would be received by the antenna from the black body is identical to the noise power received by the antenna from its actual surroundings. It is in this sense that physical significance can be attached to the effective antenna noise temperature.

If the frequency is sufficiently high, then practical temperatures may be so low that the Rayleigh-Jeans approximation is no longer valid and the expression based on Planck's equation should be used. It is still possible under these circumstances to define an effective antenna noise temperature.

However, the concept of effective antenna noise temperature, as defined for the quantum mechanical case is of limited utility. In the classical case, the total input noise power to a receiver can be specified in terms of an effective temperature which is just the sum of the effective noise temperatures of the individual sources of noise. It is this additive property that makes the concept of effective temperatures very useful in the classical case. In the quantum mechanical case the effective noise temperatures are no longer additive because the noise powers are no longer linearly related to the effective noise temperatures.

Calculation of the Antenna Noise Power for Some Practical Cases

The integration of Equation (12-4) to obtain the total available noise power is quite difficult in practice. Not only are the effective area and

brightness complicated functions of the coordinates but usually they are known only approximately, at least for certain values of θ and ϕ. In practice, Equation (12-4) is usually evaluated by the summation of a finite number of terms. Certain practical cases have been considered. The results are given below in terms of both brightness and effective brightness temperatures.

1. The solid angle subtended by the source (Ω_s) is less than the antenna beamwidth (i.e., $\Omega_s < \Omega_a$). Assume that the coordinates of the source center are θ_1 and ϕ_1.

a. Assume that the brightness of the source and the effective area of the antenna are constant over the solid angle subtended by the source (Ω_s) and that the brightness is zero elsewhere. That is $B(\theta, \phi) = B$ (a constant) within the solid angle Ω_s. $B(\theta, \phi) = O$ outside the solid angle Ω_s.

and

$$A(\theta, \phi) = A(\theta_1, \phi_1) \text{ (a constant)}$$

where

$A(\theta_1, \phi_1)$ is the effective area in the particular direction θ_1, ϕ_1.

Substituting those quantities in Equation (12-4) and integrating we have an expression for the available power.

$$P = \gamma\beta A(\theta_1, \phi_1) B\Omega_s \tag{12-16}$$

If the main beam is pointed in the direction of the source, then the effective area in the direction of the source is A and the available power is

$$P = \gamma\beta A\Omega_s \tag{12-17}$$

Expressing the effective area in terms of the beamwidth by using (12-7) and (12-8), Equation (12-17) becomes

$$P = \gamma\beta\lambda^2 B\frac{\Omega_s}{\Omega_a} \tag{12-18}$$

Expressing the constant brightness B in terms of a constant effective brightness temperature T_e, (12-18) can be written as

$$P = \gamma\beta 2kT_e \frac{\Omega_a}{\Omega_s} \tag{12-19}$$

In practice, the case of $\Omega_s < \Omega_a$ occurs for such noise sources as the sun. Of course, the antenna gain must be low enough so that the antenna beamwidth is greater than the solid angle of the sun. Actually both the brightness and effective area functions will usually vary somewhat over the solid angle subtended by the source but they can frequently be approximated by a constant representing a suitable average. This is particularly true if the source is much smaller than the antenna beamwidth.

b. A point source corresponds to the limiting case of $\Omega_s < \Omega_a$. Practical sources are radio stars, since their angular extent is extremely small. For a true point source it is not possible to specify the brightness and angular extend individually since as $\Omega_s \to 0$ $B \to \infty$. Hence, for point sources, the flux density $S = \iint_{\Omega_s} B(\theta, \phi)\, d\Omega$ (watts $m^{-2}(cps)^{-1}$) is usually given. Then

$$P = \gamma\beta A(\theta, \phi)S \tag{12-20}$$

2. The solid angle subtended by the source is greater than the beamwidth ($\Omega_s < \Omega_a$).

a. The brightness is constant over the whole solid angle of 4π steradians. That is $B(\theta, \phi) = B$ (a constant) for all θ and ϕ. As already mentioned this is the situation that prevails inside a black body cavity provided $\gamma = 1/2$. In general for any γ and with constant brightness, Equation (12-4) for the available power yields

$$P = \gamma\beta B \iint_{4\pi} A_e(\theta, \phi)\, d\Omega \tag{12-21}$$

Now the integral of the effective area over the whole solid angle is just λ^2. Therefore,

$$P = \gamma\beta\lambda^2 B \tag{12-22}$$

This can again be written in terms of the effective brightness temperature using (12-2). Hence, the available power for this case in terms of brightness temperature is given by

$$P = \gamma 2kT_e B \qquad (12\text{-}23)$$

By setting $\Omega_s = \Omega_a$, this equation can also be obtained from Equation (12-19) which is based on the concept of beamwidth. It is therefore apparent that although the concept of beamwidth is an approximation, the correct result is still obtained for the particular case of constant brightness. Equation (12-23) is also a good approximation if the brightness is constant over the beamwidth of the antenna and the contribution from the side lobes is small.

12.3 COSMIC BACKGROUND RADIATION

12.3.1 Diffuse Background

A survey of diffuse radio sources in our galaxy has been made by Ko[2]. Figures 12-1 through 12-8 summarize the most reliable measurements of galactic noise spatial distributions over the frequency range 60 mcps to 910 mcps. The isophotes give sky brightness, temperatures in degrees Kelvin and are plotted in celestial coordinates Epoch 1950. The relative accuracy within each map is stated to be 5 percent to 20 percent.

The maps have general similarities but differ in local structure. The general structure of the isophotes may be conveniently represented as a superposition of symmetrical and non-symmetrical distributions. There are three symmetrical distributions:

1. A narrow bright belt of radiation about 3 degrees wide lies in the galactic plane and is concentrated toward the galactic center. This belt dominates entirely above 400 mc.
2. A very broad band of radiation is concentrated within about 15 to 30 degrees of galactic latitude and has its maximum brightness in the region of the galactic center. The brightness

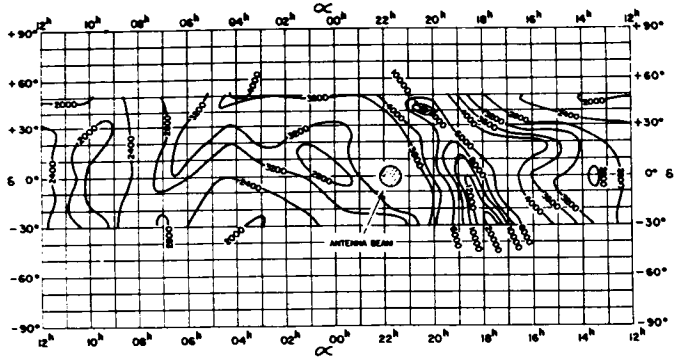


Figure 12-1. Map of the radio sky background at 64 mc (after Hey, Parsons, and Phillips). The contours give the absolute brightness temperature of the radio sky in degrees Kelvin.

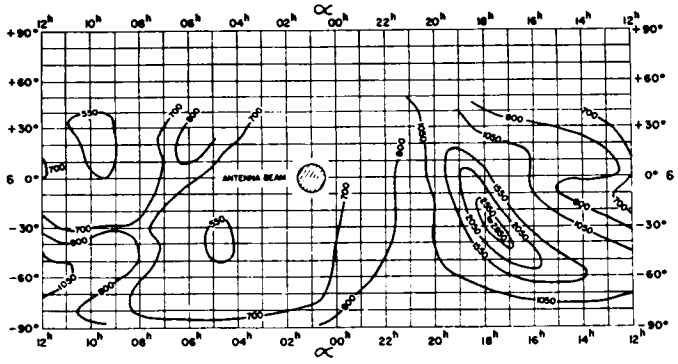


Figure 12-3. Map of the radio sky background at 100 mc (after Bolton and Westfold). The contours give the absolute brightness temperature of the radio sky in degrees Kelvin.

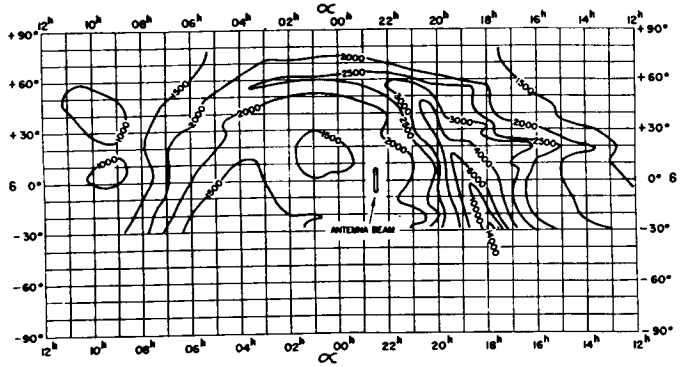


Figure 12-2. Map of the radio sky background at 81 mc (after Baldwin). The contours give the absolute brightness temperature of the radio sky in degrees Kelvin.

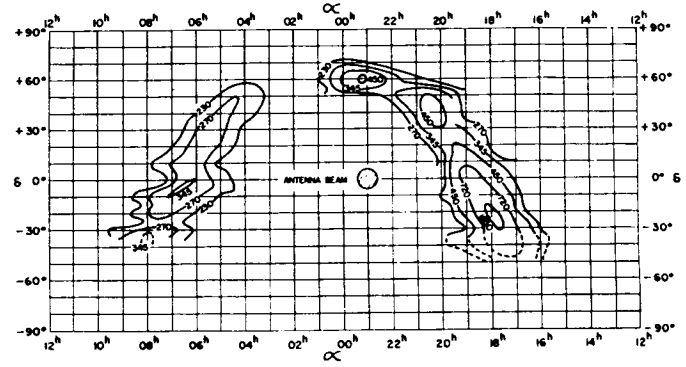


Figure 12-4. Map of the radio sky background at 160 mc (after Reber). The contours give the absolute brightness temperature of the radio sky in degrees Kelvin.

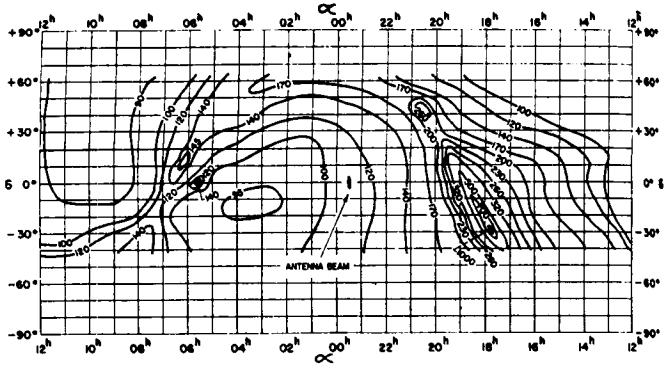


Figure 12-5. Map of the radio sky background at 250 mc (after Ko and Kraus). The contours give the absolute brightness temperature of the radio sky in degrees Kelvin.

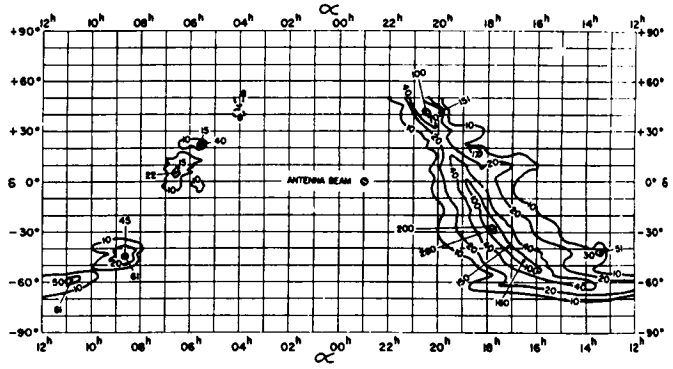


Figure 12-7. Map of the radio sky background at 600 mc (after Piddington and Trent). The contours give the absolute brightness temperature of the radio sky in degrees Kelvin.

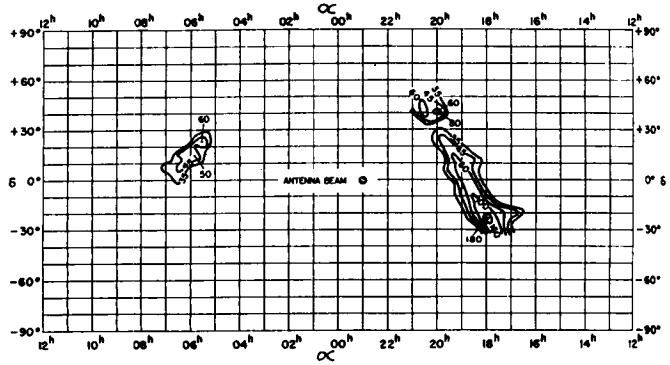


Figure 12-6. Map of the radio sky background at 480 mc (after Reber). The contours give the absolute brightness temperature of the radio sky in degrees Kelvin.

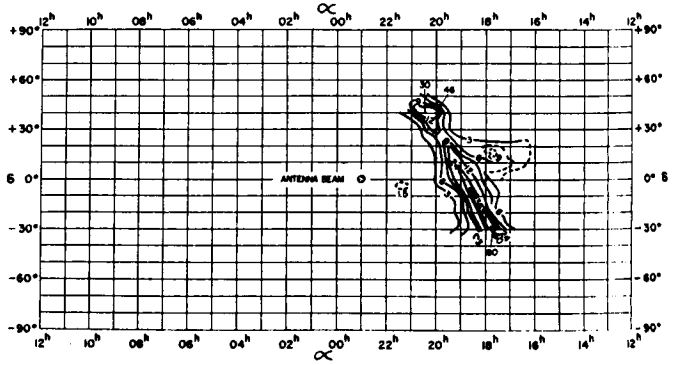


Figure 12-8. Map of the radio sky background at 910 mc (after Denisse, Leroux, and Steinberg). The contours give the absolute brightness temperature of the radio sky in degrees Kelvin.

is an inverse function of frequency. This band dominates at frequencies below 250 mc and decreases to insignificance above 400 mc.

3. There is a roughly isotropic component of radiation distributed over the entire celestial sphere. The brightness of this component varies inversely with frequency. Above 200 mc, the brightness is so low that measurement becomes difficult.

Above 1 gc the background galactic brightness temperature is very low and exclusively concentrated in the galactic center. Brown and Hazard[3] quote 17°K at 1.2 gc with a 2.8° beam and 2.6°K at 3 gc with a 3.4-degree beam.

The data presented in Figures 12-1 through 12-8 as discussed above is summarized in Figure 12-9[4]. Figure 12-9 depicts noise temperature (°K) versus frequency for the galactic center and for one of the coldest regions of the radio sky. Space communication systems can expect noise temperatures near the minimum value for most regions of the sky. It is apparent from Figures 12-1 through 12-9 that the average galactic background temperature varies inversely with frequency. Several investigators[5, 6] have concluded that the brightness spectrum is of the form

$$B \propto \lambda^n \tag{12-24}$$

where n is approximately 2.5 over most of the spectrum.

Of particular interest in space communications is the level of galactic noise along the ecliptic plane. The majority of more distant space missions will presumably occur in or near the ecliptic plane and hence it is the ecliptic noise level which will be present as an interfering background. Figure 12-10[7] shows curves of the ecliptic noise level for three different frequencies, 81 mcps, 100 mcps, and 600 mcps. Extrapolations to higher frequencies may be made by means of the aforementioned frequency dependence of equation (12-24).

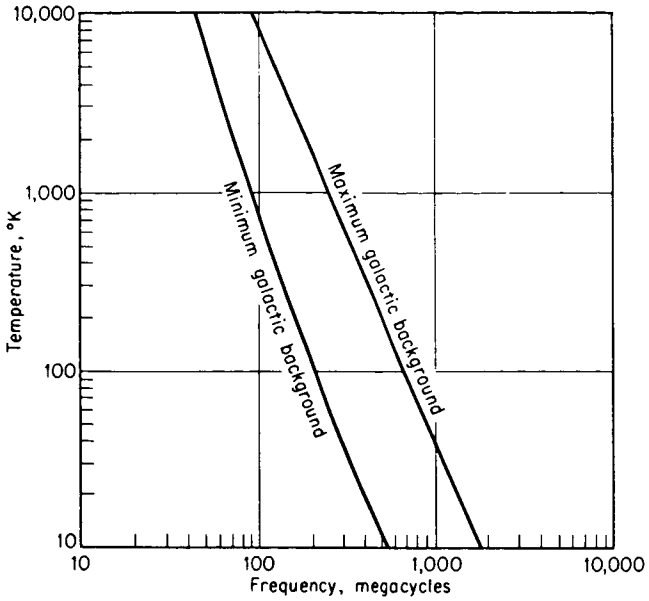


Figure 12-9. Galactic noise temperature.

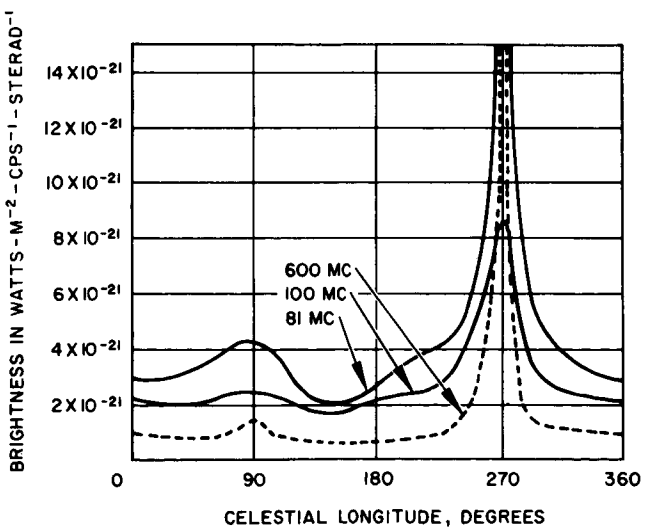


Figure 12-10. Galactic noise along the ecliptic at three frequencies. The 600-mc peak extends to a height of 21 units and the 81-mc peak to 27 units.

12.3.2 Radio Stars

Superimposed on the general galactic background radiation are numerous discrete sources, generally less than 1 degree in extent. Since the majority of these sources cannot be identified with visible objects they are known as radio stars. The strongest of these sources tend to occur near the galactic plane. Figure 12-11[8] locates the ten strongest radio stars in equatorial coordinates and gives their noise temperatures at 378 mc. Figure 12-12[9] for comparison purposes locates most of these same sources in celestial coordinates on the galactic noise map of Figure 12-5.

The spectral irradiance of the four brightest radio stars outside the terrestrial atmosphere in terms of flux density is shown in Figure 12-13[10]. Spectra of radio stars have been found to fall into one of three categories. Two categories, Class S and Class C, follow a power law relation of the form

$$\text{Log } B = -\alpha \log f + \text{constant} \qquad (12\text{-}25)$$

where α is called the spectral index. Class S sources have constant spectral indices within experimental accuracy below 1420 mc. Class C sources have spectral indices which vary with frequency. Class T sources have a thermal spectrum. Of 158 sources for which information is available 134 are Class S, 19 are Class C, and the remaining 5 are Class T. Spectra have been tabulated for all 158 sources by Conway, Kellerman and Long[11].

For a receiving system in the 10 g cps to 300 gcps region, the contribution of radio stars to the total noise power can in general be neglected. Not only is the contribution small when they are directly in the main beam, but because of their discrete nature they can easily be avoided.

12.3.3 Hydrogen Line Emission by Interstellar Gas Clouds

Radio astronomers have detected an essentially monochromatic spectrum line due to the radiation of atomic hydrogen at a wavelength

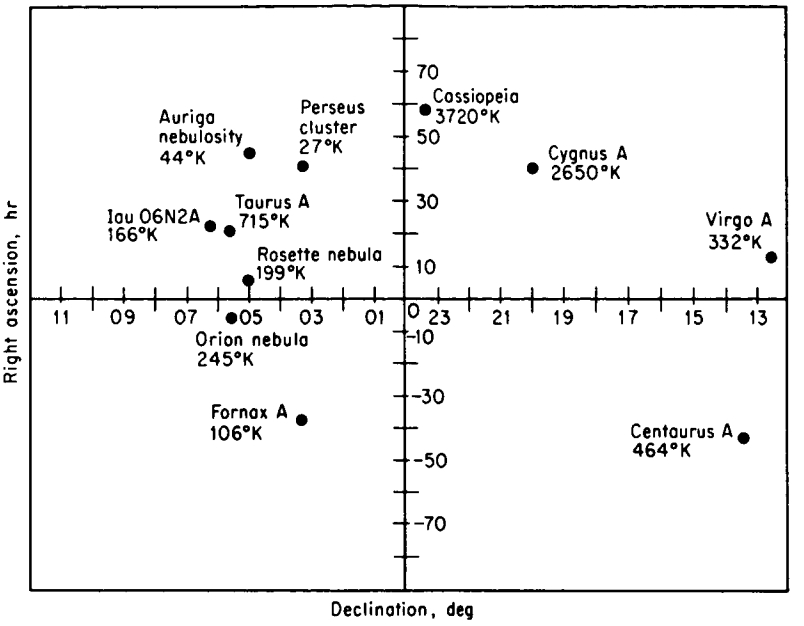


Figure 12-11. Radio stars at 378 mc.

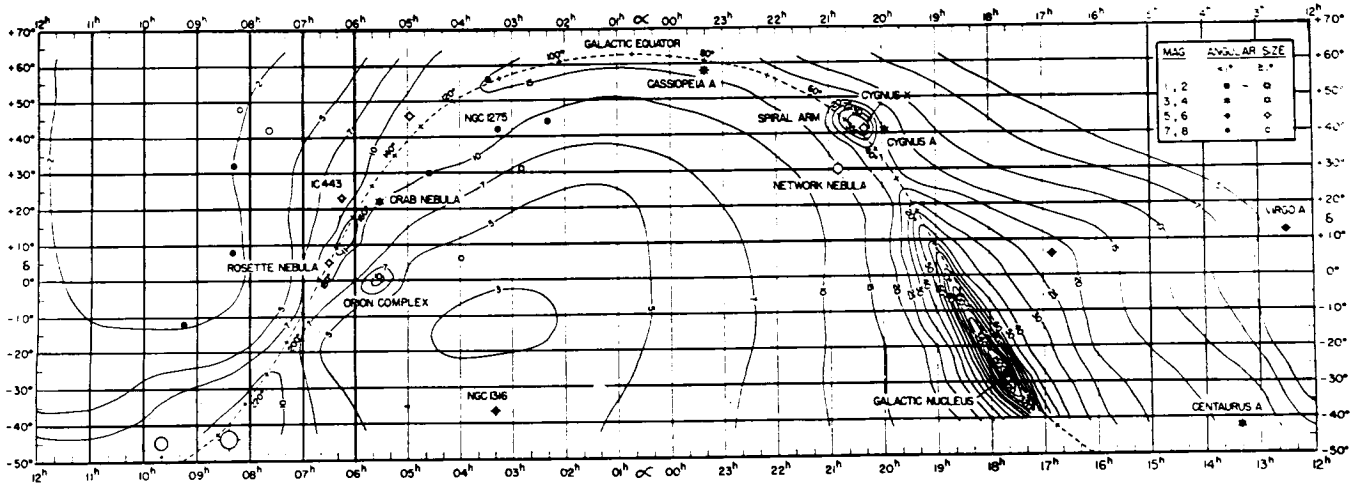


Figure 12-12. The most prominent radio stars superimposed on the radio sky background map at 250 mc. (Brightness temperatures of the isophotes are in degrees Kelvin.)

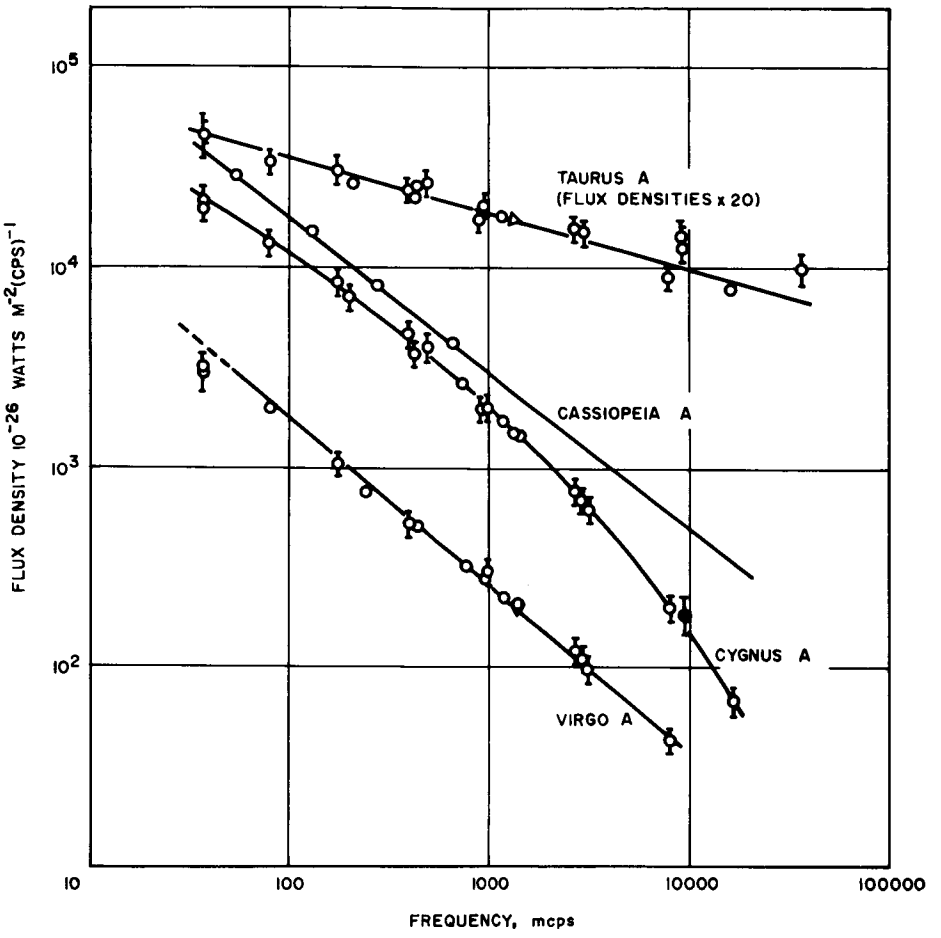


Figure 12-13. Spectra of the most prominent radio stars.

of 21 cm (1.42 gcps). The hydrogen line emission is a maximum along the Milky Way with a distribution over the sky roughly similar to the general galactic background. It has a maximum brightness temperature of 100^{o} and maximum brightness of 6×10^{-20} watts/m^2 steradian per cps[12]. This is well in excess of the general galactic background at that frequency. However, since this is limited to a single wavelength, it is unimportant as a source of interference.

12.3.4 Solar System Background

1. Solar Radiation

The sun is an extremely important noise source for certain frequencies. For example, Mercury is never more than 28^{o} from the sun when viewed from the earth, while earth as viewed from Jupiter is always within 11^{o} of the sun.

The r-f radiation from the sun can be classified into five categories:

a. The thermal or basic component.
b. The slowly varying component which is associated with sunspots and has a period of approximately 27 days (the mean solar rotation period).
c. Noise storms, consisting of trains of bursts together with enhancement of the general radiation; these last for hours or days and show strong circular polarization.
d. Outbursts associated with flares which can be very intense and last for a number of minutes.
e. Isolated bursts lasting for 5 to 10 seconds.

In the frequency range, 30 to 300 g cps, the sun radiates essentially as a black body (thermal component) at a temperature of 6000^{o} to 7000^{o}K. There is some slight enhancement due to the solar flares.

In the 0.3 to 30 g cps frequency range the solar radiation is more intense than for a black body at 6000 to 7000^{o}K. In addition, the radiation exhibits temporal effects as the contribution of sunspots and

flares (particularly sunspots) becomes more important; at 10 g cps enhancements of the order of 2 occur. This is a small enhancement compared with enhancements of the order of 10^3 at meter wavelengths[13]. The observed values of sun temperature between 0.3 g cps and 35 g cps follow closely the relationship

$$\frac{T_s}{290} = \frac{675}{f}\left[1 + \frac{1}{2.3}\sin 2\pi \frac{\log_{10} 6\,(f - 0.1)}{2.3}\right] \qquad (12\text{-}26)$$

where T_s is the apparent sun temperature and f is the frequency in gcps. Equation (12-26) is plotted in Figure 12-14[14].

From earth, the sun subtends a solid angle, Ω_s, of about 7.6×10^{-5} steradians. The temperature given by the above equation will be observed by an antenna having a beamwidth, Ω_a, less than this. For antennas having beamwidth greater than this, the apparent temperature is decreased by the ratio of the sun angular substense to the antenna beamwidth. Measured solar irradiance levels outside the earth's atmosphere are depicted in Figure 12-19[15].

2. Planetary Radiation

For low noise space receiving systems in the vicinity of a planet, the planetary radiation must be taken into consideration. However, to an earth based receiving system, the planets contribute negligible noise power even if the planet is in the main beam of the antenna. This is because the solid angles subtended by the planets at earth are so small that they fill only a small part of the antenna beam.

Thermal radiation from Venus, Mars and Jupiter has been measured by Mayer and others[16, 17] (Table 12-1).

Figure 12-15[16] shows the theoretical maximum thermal radiation outside the earth's atmosphere based on planetary temperature estimates. These are essential in agreement with infrared measurements except in the cases of Jupiter and Venus.

The discrepancy between the microwave and infrared temperatures of Venus is customarily explained by assuming that microwaves

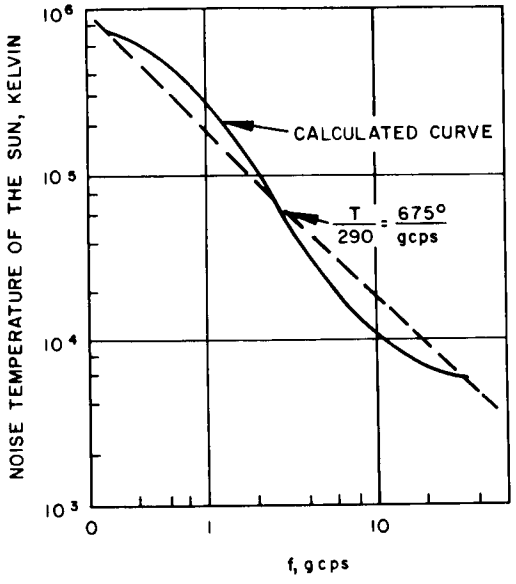


Figure 12-14. Sun noise temperature versus frequency.

Planet	Infrared Temperature (^{o}K)	f(gcps)	λ(cm)	Equivalent Radio Blackbody Temperature (^{o}K)
Mercury			4	400^{o}K (Planet at greatest elongation)
Venus	240	35	0.86	410^{o}K
			2-21	620^{o}K
Mars	260	9.5	3.15	211^{o}K
Jupiter	130		3-8	145^{o}K
		2.9	10	600^{o}K
		1.4	21	2000^{o}K
		0.97	31	5000^{o}K
		0.94	136	50000^{o}K
Saturn	150	8.7	3.45	106^{o}K
		2.9	10	140^{o}K
			Highly polarized	200^{o}K

Table 12-1. Planetary Radiation.*

*Mayer, et. al.[16, 17]

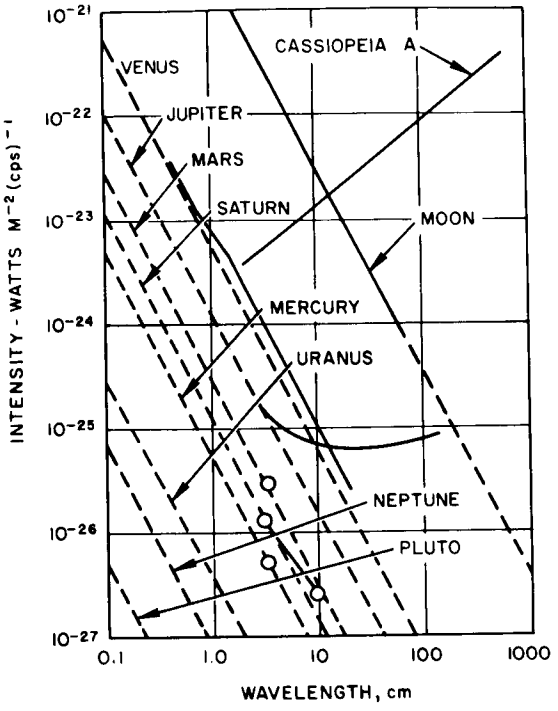


Figure 12-15. The predicted intensity of planetary thermal radio radiation when the planet is near closest approach to the earth (dashed lines), and the measured intensity of radiation from the planets, the moon, and the nonthermal radio source, Cassiopeia A (solid lines and points).

penetrate the atmosphere of the planet more readily than infrared rays. Hence, the infrared temperature is the temperature of the atmosphere whereas the microwave temperature is that of the actual surface. In the case of Jupiter the discrepancy is believed due to radiation by synchrotron radiation by high energy electrons trapped in the planetary magnetic field. This theory is supported by observations by Radhakreshnan and Roberts of CIT that the signals are linearly polarized and come from an area larger than the solid disc of the planet.

In addition to the above mentioned steady state thermal radiation, several planets notably Jupiter, Saturn and Venus exhibit strong sparodic bursts of clearly non-thermal radiation; typical peak intensities observed to date include 10^{-19} watt m^{-2} $(cps)^{-1}$ at 19.6 mcps (Jupiter), 6×10^{-21} watt m^{-2} $(cps)^{-1}$ at 22 mcps (Saturn), and 9×10^{-22} watt m^{-2} $(cps)^{-1}$ at 26.7 mcps (Venus)[18]. In no instance have these impulsive signals been observed at frequencies above 43 mcps.

3. Lunar Radiation

The thermal noise flux from the moon is considerably greater than that from the planets. Furthermore, the moon subtends an angle of approximately 0.5 degrees at the earth, and hence, practical antennas in the mm region can have beamwidths less than the solid angle subtended by the moon (6×10^{-5} steradians). This means that provided the earth's atmosphere permits, such an antenna based on earth can "see" the moon directly. Thus for certain frequencies in the 10-300 g cps region, where atmospheric effects are smallest, the moon can make a significant contribution to the noise temperature of an antenna pointed at the moon. Gibson has found that at 35 g cps the brightness temperature of the moon varies with phase between the limits of 145°K and 220°K as shown in Figure 12-16[19]. The corresponding flux densities at earth are 4×10^{-21} watts m^{-2} $(cps)^{-1}$. The brightness temperatures determined by Gibson should be reasonably valid throughout the whole frequency range of 10-300 g cps, and hence can be used to calculate the noise contribution from the moon for any particular case. Lunar surface temperatures as determined by infrared measurements vary from

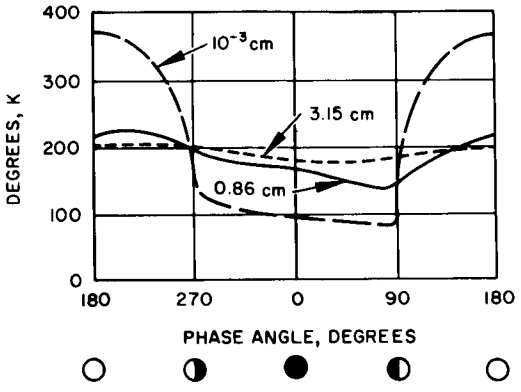


Figure 12-16. The brightness temperature of the center of the disk of the moon as a function of the phase of solar illumination. Long dashes - brightness temperature corresponding to the infrared radiation as approximated by the calculated surface temperature of Wesselink (Bull. Astron, Inst. Neth., vol. 10, pp. 351-363; April 1948). Solid line - brightness temperature corresponding to the 8.6-mm radiation measured by Gibson (Proc. IRE, vol. 46, pp. 280-286; January 1958). Short dashes - brightness temperature corresponding to the 3.15 cm radiation measured by Mayer, McCullough, and Sloanaker "Planets and Satellites," University of Chicago Press, Chicago, Ill., Ch. 12; 1961). Reproduced from "Planets and Satellites," G. P. Kuiper and B. M. Middlehurst, Eds., University of Chicago Press, Chicago, Ill., 1961; copyright 1961 by the University of Chicago.

120°K to 400°K and are observed to lag the microwave temperatures by approximately 45 degrees. This feature may be accounted for by postulating a thin layer of lunar surface dust which is transparent to microwaves but effectively insulates an underlying rock surface. At frequencies less than 1.4 g cps the radiation temperature is a constant 250°K[20].

12.3.5 Terrestrial Atmospheric Radiation

The earth's atmosphere radiates thermal noise as a result of absorption of incident radiation. The effect of ionospheric radiation can be approximated by assuming that each 0.1 db of absorption is equivalent to 7°K antenna noise temperature. Since the ionospheric absorption is less than 0.1 db for frequencies greater than 0.3 g cps ionospheric absorption noise can be neglected in this regime[21].

Tropospheric noise is primarily due to absorption by oxygen and water vapor. Thus the antenna is immersed in an atmosphere that emits black body radiation. The tropospheric contribution to noise temperature of a narrow beam antenna having a radiation pattern which admits no side or back lobes is

$$T_s = \int_0^\infty \alpha T \exp\left[-\int_0^r \alpha \, dr\right] dr \qquad (12\text{-}27)$$

where α and T are respectively the absorption coefficient and temperature at a distance r from the antenna. Calculated curves of T_s versus frequency are shown in Figure 12-17[22] for various values of antenna beam zenith angle. As expected the temperature curves show maxima at the absorption peaks of water vapor and oxygen. Also, at any frequency the sky noise varies inversely with the elevation angle. This is due to the longer path length required to fully penetrate the atmosphere at low elevation angle.

12.4 OPTICAL BACKGROUND

12.4.1 Cosmic Background Radiation

Ramsey[23] has prepared charts giving spectral irradiance outside the earth's atmosphere of the brightest stars. In making these calculations it was assumed that the stellar spectra follow Planck's law. Figure 12-18 depicts spectral irradiance reaching the top of the earth's atmosphere from the stars which exhibit the greatest irradiance in the visible region. Variable stars such as Betelguex, Mira, and R. Hydrae are presented at their maximum irradiance.

12.4.2 Solar System Background

1. Solar Background Radiation

The mean solar irradiance over all wavelengths outside the earth's atmosphere at the mean solar distance is called the solar constant is approximately 140 mw/cm^2 or 130 watts/ft^2. Over 98 percent of this solar irradiance is contained in the wavelength region 0.3 to 4.0 microns. Variation in solar distance throughout the year causes this irradiance to vary as much as 3.5 percent from the mean; solar activity can produce fluctuations of 1.5 percent in the solar constant. In addition to these relatively small variations, the amount of solar irradiation that reaches a particular portion of the earth's surface varies with solar elevation angle and atmospheric conditions.

The spectral distribution of solar radiation in the visible and near infrared resembles that of a black body at 6000°K. H. Malitson[15] has compiled a chart of all available measurements of solar spectral irradiance in the wavelength range 1 Å to 10 m. Figure 12-19 shows the spectral distribution of solar radiation outside the atmosphere at the earth's mean solar distance.

2. Planetary and Lunar Radiation

The steady state background irradiance from the planets and the moon results from two processes: reflection of incident solar energy

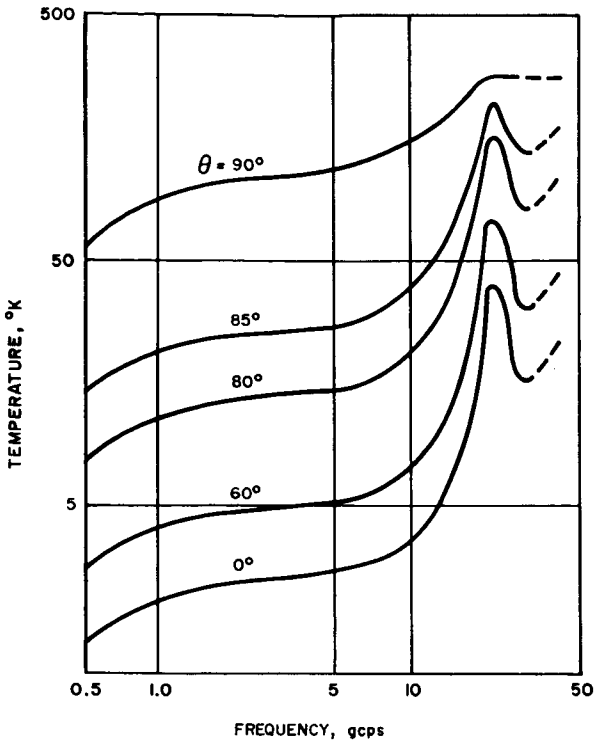


Figure 12-17. Sky noise temperature due to oxygen and water vapor.

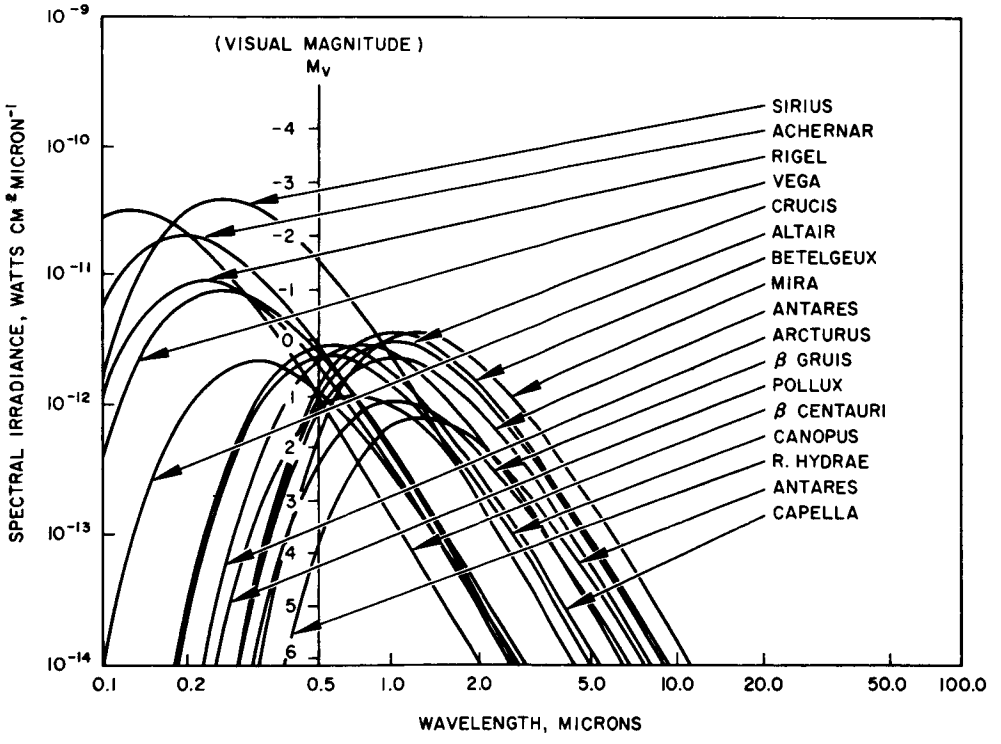


Figure 12-18. Spectral irradiance of brightest stars outside the terrestrial atmosphere.

and self-emission as a result of the planets non-zero temperatures. Figures 12-20[24] and 12-21[25] show the background spectral irradiance outside the terrestrial atmosphere due to reflected and self-emitted energy for the planets and mean respectively. These curves assume constant the albedo at peak power wavelength. The planetary irradiances at other than peak power wavelengths may be determined from these curves and the spectral albedo which is plotted in Figure 12-22[26]. The peak spectral irradiances due to self-emitted radiation were calculated from

$$H_{peak} = W_{max} \frac{(1-A)}{\pi} \frac{\pi d^2}{4R^2} \qquad (12\text{-}28)$$

where

H_{peak} = peak spectral irradiance (watts $cm^{-2}\ \mu^{-1}$)

A = bond albedo

W_{max} = peak spectral radiant emittance at wavelength λmax

where λmax is given by Wien's law

$$\lambda max = \frac{2898}{T^{(^oK)}} \qquad (12\text{-}29)$$

These data are summarized on Table 12-2. The values of T_e for each planet were obtained from Kuiper[27]. The peak irradiances from reflected solar radiation are based on the solar illumination curves for a 6000°K black body and the planetary albedo at the peak power wavelength. The superior planets (those outside the earth's orbit) are assumed to be at oppostion, the inferior planets (those within the earth's orbit) are assumed to be at greatest elongation. These astronomical terms are explained by Figure 12-23.

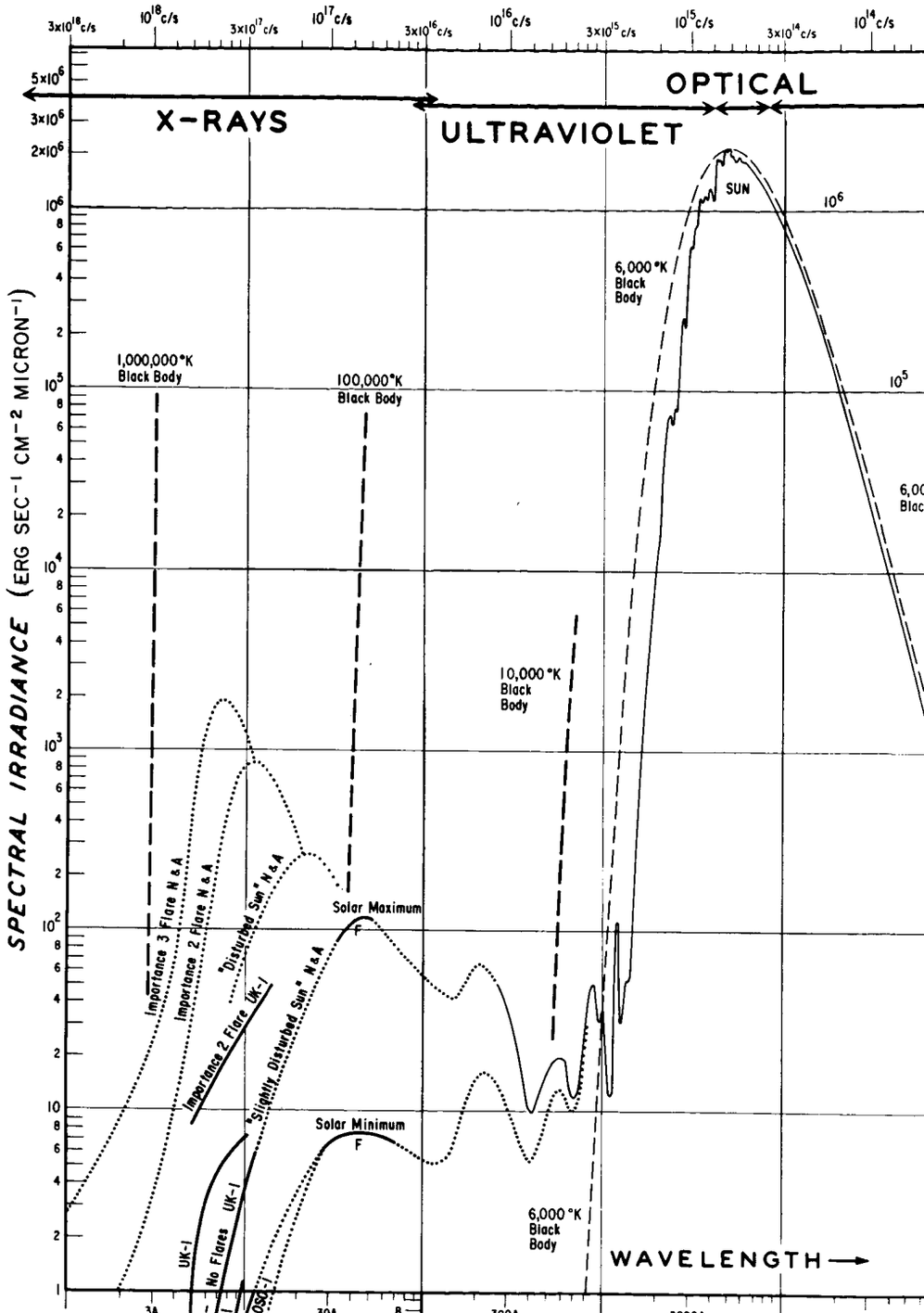
THE SOLAR SPECTRUM

Displayed here is the entire radiation pattern of the sun, from X-rays through visible light to radio waves. The solar energy received at the top of the earth's atmosphere can be read from the vertical scales. Wavelengths to which these energies apply are indicated at the bottom, frequencies at the top.

The units of spectral irradiance read, in unabbreviated form, "ergs per square centimeter per second in a wavelength interval of one micron." Because of the enormous range of energies charted here, it was necessary to fold this logarithmic scale thrice in the infrared and radio regions. The folded sections are accompanied by the appropriate numerical labels.

Along the scale at the top, frequencies in cycles per second (c/s) increase toward the right. One gigacycle per second (Gc/s) is equivalent to 10^9 (one billion) cycles per second; therefore, 1,000 gigacycles equals 10^{12} cycles. Farther to the right, one megacycle per second (Mc/s) is 10^6 (one million) cycles per second.

The wavelengths, given along the bottom scale, become longer toward the right; they are indicated successively in angstroms (A), microns (μ), millimeters (mm), centimeters (cm), and meters (m). One micron is 10,000 angstroms, and one millimeter is 1,000 microns.



F = Friedman, H., "Solar Radiation", Astronautics, 14-23, August 1962 (rocket measurements).

M & A = Nicolet, M and Aikin, A.C., "The Formation of the D Region in the Ionosphere", J. Geophys Res. 65, 1469-1483, 1960.

OSO-1 = Data from the first Orbiting Solar Observatory W A White, private communication, 1964.

UK-1 = Pounds, K A., Willmore, A P, Bowen, P J., Norman, K., and Sanford, P W., "Measurements of the Solar Spectrum in the Wavelength Band 4-14A". Published in Proc Roy. Soc., 1964. Ariel I satellite data.

Measurements by Hinteregger (1) in Litter, "Space Astrophysics", p 81, 1961. (2) "Analysis of the Solar Emission Spectrum from 1300 to 250A as observed in Aug. 1961" (with Zinn & Hall), 3rd COSPAR meeting, 1962. Estimates based on above and on OSO-1 data in preprint X-615-64-162 (Goddard Space Flight Center) by Bourdeau, R E., Chandra, S., and Neupert, W M., June 1964

Detwiler, Garrett, Purcell, Tousey, "The Intensity Distribution in the Ultraviolet Solar Spectrum," Ann de Geophys. 17(3), 263-272, 1961.

Johnson, F.S., "The Solar Constant", J. Meteor. 11(6), 431-439, 1954 (Includes work of Dunkelman and Scolnik, J. Opt. Soc. Am. 49(4), 356-367, 1959).

A

12-29

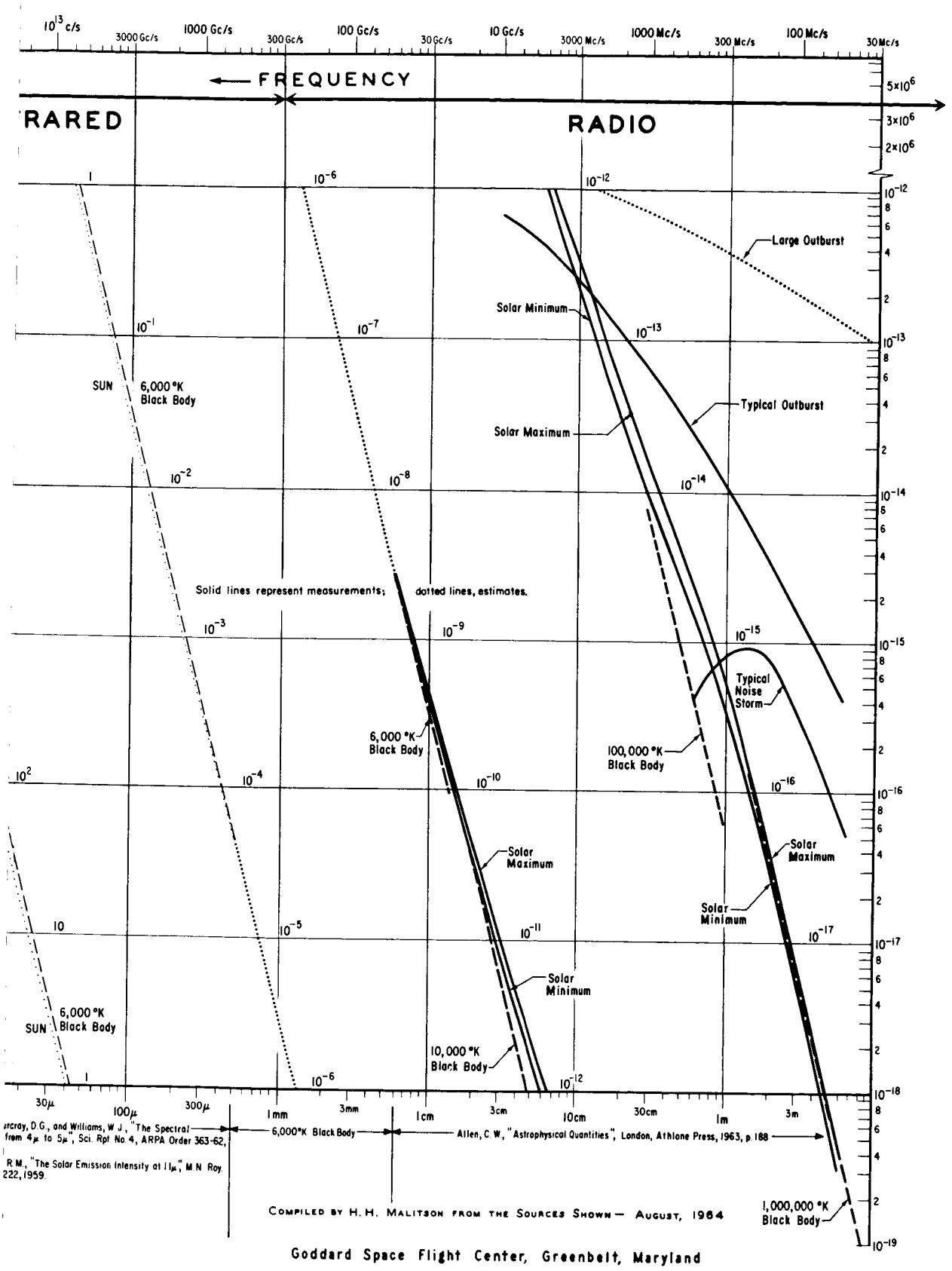


Figure 12-19. Solar spectral irradiance outside the earth's atmosphere.

Planets	Distance to Earth (times 1×10^{16} km)	Diameter (times 1000 km)	Bond Albedo	Temperature (°K)	Peak Wavelength (μ)	Max Mission at Peak Wavelength $\left(\frac{\text{watt}}{\text{cm}^2\mu}\right)$	Visual Magnitude	Angle to Eliptic Plane (Degrees)
Mercury	137.8	4.8	0.056	613	4.65	3.36×10^{-11}	+ 0.11	7
Venus	103.3	12.4	0.76	235	12.2	8×10^{-13}	- 4.23	3.3
Moon (satellite)	0.384	3.45	0.067	373	7.7	3×10^{-7}	-12.54	–
Mars	78.3*	6.8	0.16	217	10	1.02×10^{-12}	- 2.01	1.8°
Jupiter	628.3	142.8	0.73	138	20.7	2.24×10^{-13}	- 2.55	1.3°
Saturn	1277	120.8	0.76	123	23.3	1.85×10^{-14}	+ 0.67	2.5°
Uranus	2720	47.6	0.93	90	32μ	4.8×10^{-17}	+ 5.8	0.66°
Neptune	4330	44.6	0.84	–	–	–	+ 7.6	1.7°
Pluto	11, 200	14.4**	0.14	–	–	–	+14.7	17.1°***

*Distance between Mars and earth varies because of the orbit of Mars. At favorable opposition which occurs every 15 or 17 years during Aug - Sept, the distance can be as small as 55×10^6 km whereas at unfavorable opposition the distance is 105×10^6 km.

**The value measured by Kuiper (26) is believed to be much too small. When this value of diameter is used, the density of Pluto comes out to be unrealistically high.

***Pluto's orbit is quite eliptical and a portion lies inside the orbit of Neptune however they do not intersect.

The Bond albedos and planet temperatures were taken from Reference 2-6. Visual magnitudes were calculated with the absolute visibility V(1, 0) taken from the same references. There is a wide variation in reported planetary temperatures in the literature. Variations can be as large as 60°K in the case of Venus.

Table 12-2. Summary of peak spectral irradiance data.

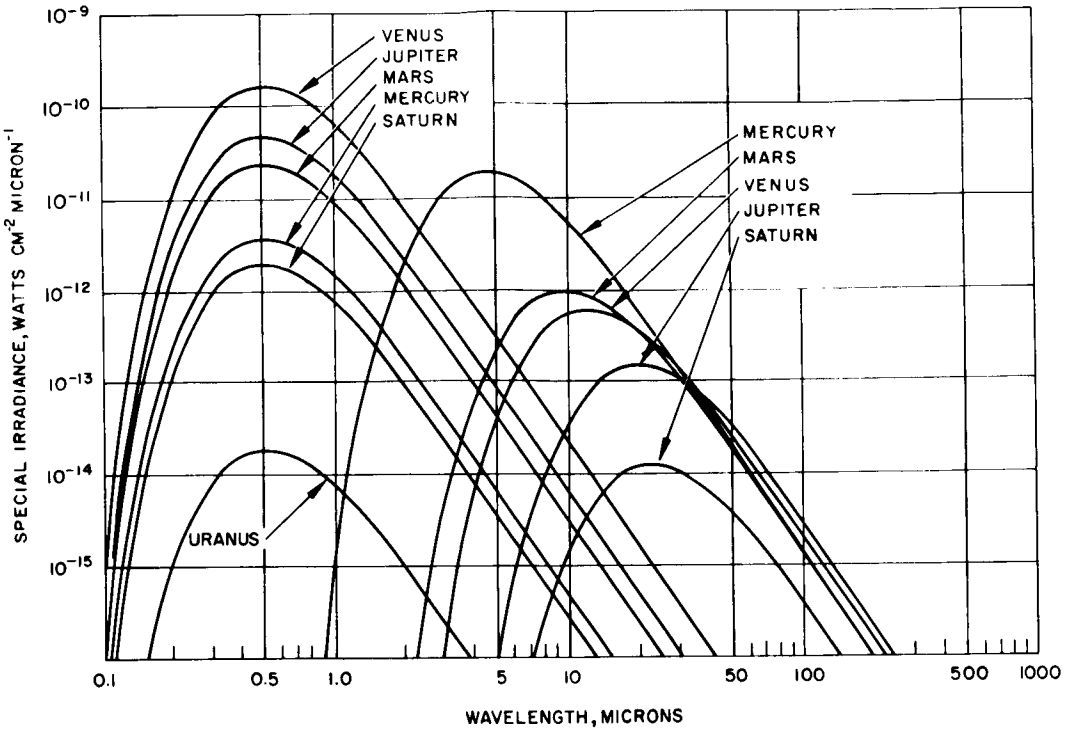


Figure 12-20. Calculated planetary special irradiance outside the terrestrial atmosphere.

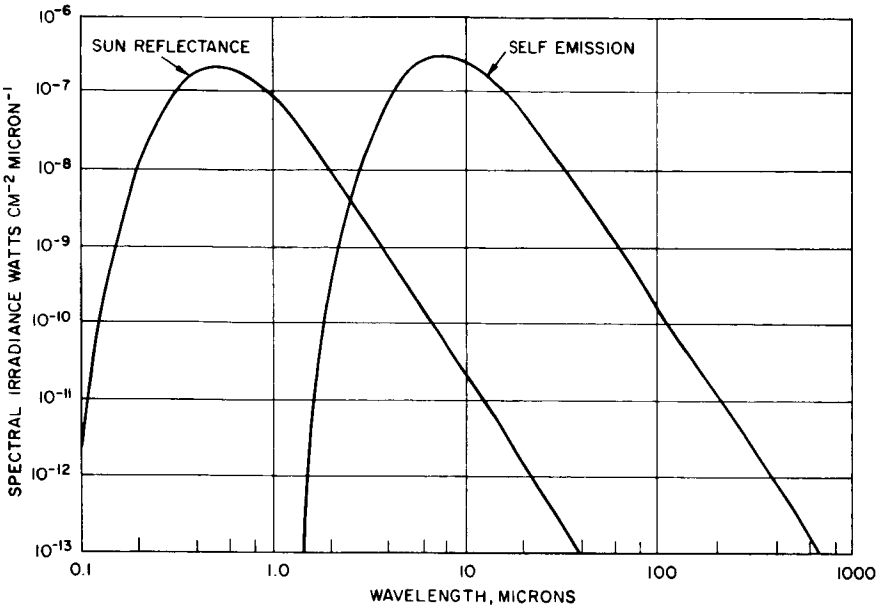


Figure 12-21. Full moon spectral irradiance outside the terrestrial atmosphere.

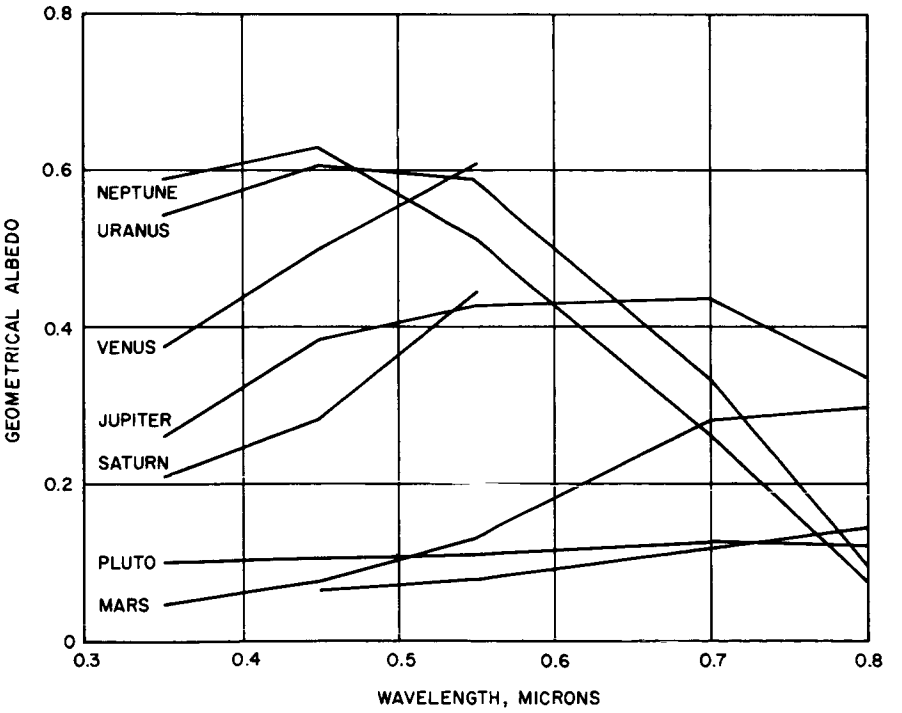


Figure 12-22. Spectral albedo versus wavelength.

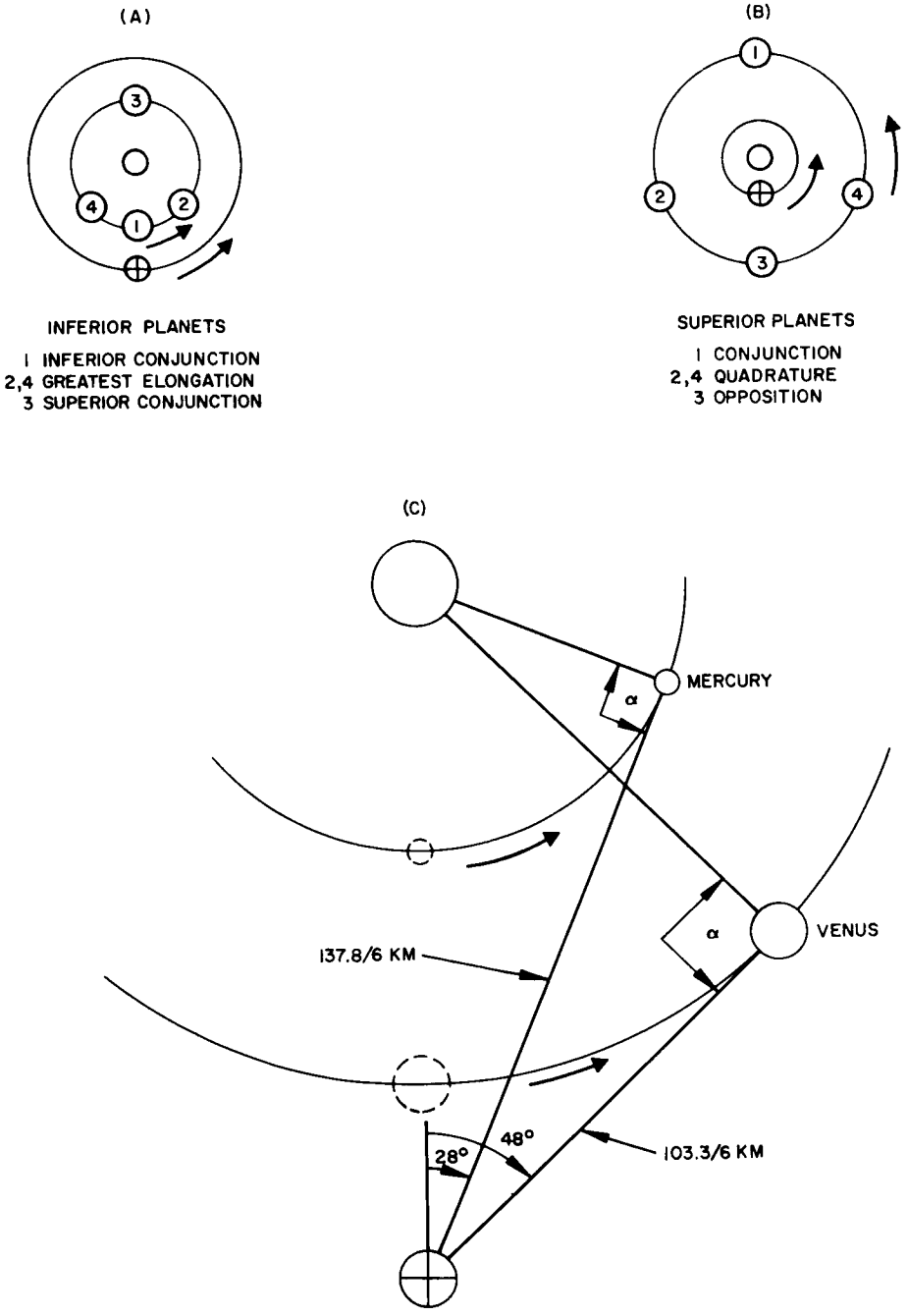


Figure 12-23. Position of Mercury and Venus at greatest elongation ($\alpha = 90^{\circ}$).

Earth radiation to space consists of reflected solar energy and self-emitted energy. Spherical albedos of the earth in the wavelength regions of interest are as follows[28]:

Wavelength	Spherical Albedo
Ultraviolet	0.50
Visible	0.40
Infrared	0.28

The mean albedo over the solar spectrum is approximately 0.35. In addition, the earth radiates by self-emission as a black body at a temperature of 220°K to 320°K, depending on latitude. When the albedo and black body radiation from the earth's surface is combined with the effect of selective atmospheric absorption, the spectral radiant emittance of Figure 12-24 results.

For distances sufficiently great that the earth may be treated as a point source, the spectral irradiance may be calculated by assuming that the earth is a Lambertian radiator and using spectral radiant emittance of Figure 12-24. Alternately the total irradiance over the solar spectrum may be calculated by using the solar constant and the average albedo. The spectral irradiance at a distance R from the center of the earth under the point source assumption is

$$H_\lambda = \frac{W_\lambda}{\pi} \frac{\pi d^2}{4R^2} \tag{12-30}$$

where

W_λ = spectral radiant emittance $\left(\frac{\text{watts}}{\text{cm}^2 \text{ micron}}\right)$ of Figure 12-24

d = diameter of the earth 12.73×10^3 km

R = distance from the earth's center (km)

In the case of reflected radiation, the above expression represents a maximum H_λ corresponding to the fully illuminated earth being viewed by the receiving surface. The problem of determining the irradiance

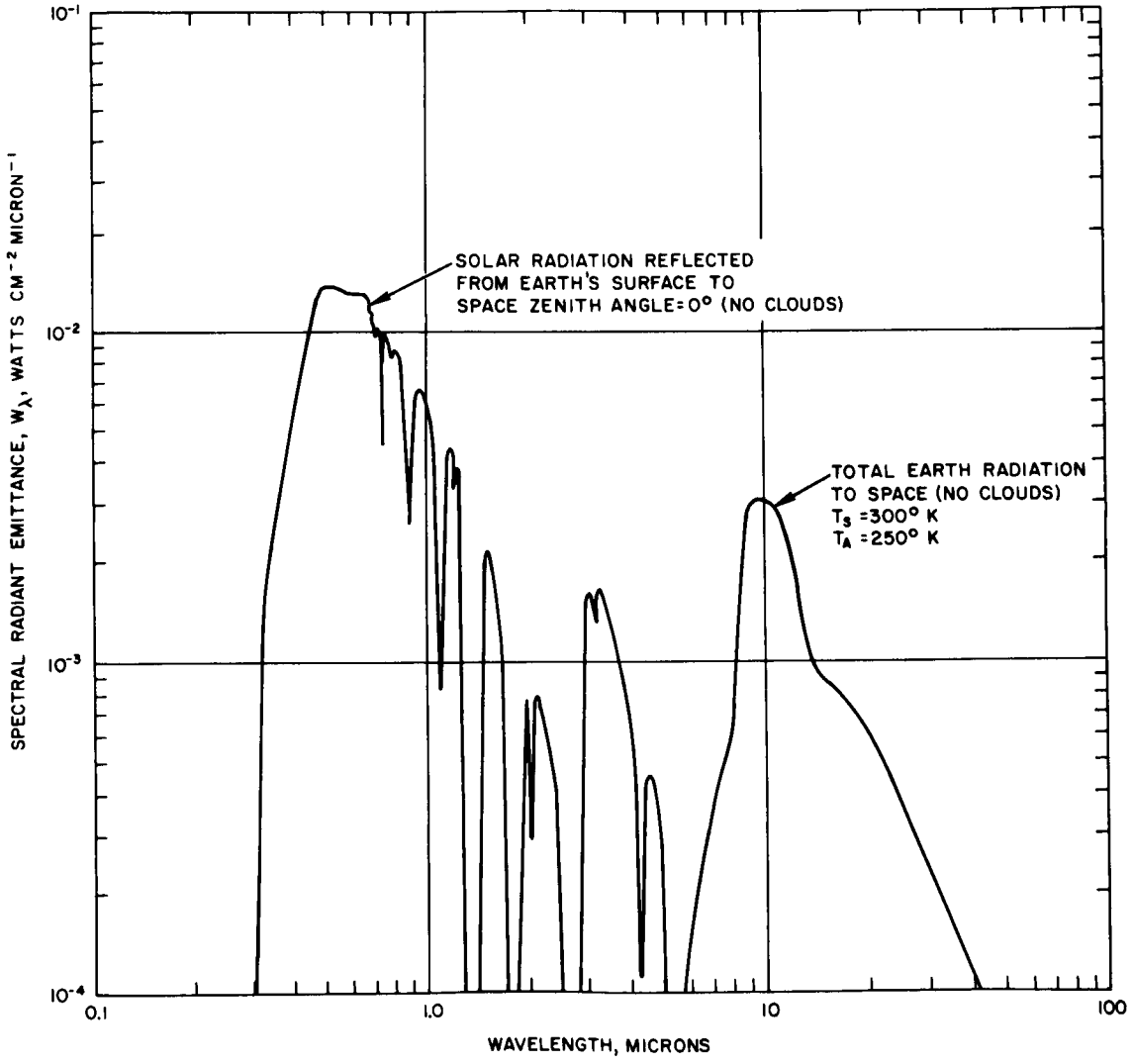


Figure 12-24. Solar and terrestrial radiation. Reflected solar and total earth radiation to space values should be divided by π to obtain the radiance for each case. T_S is the surface temperature and T_A is the effective radiating air temperature.

of a body near the earth (or any other planet) is more difficult since the view factor must be taken into account. This problem has been exhaustively treated by Dennison[29, 30]. The reflected solar irradiance H_λ incident on a unit surface of a body near a planet is given by

$$H_\lambda = W_\lambda F \tag{12-31}$$

where W_λ is the radiant emittance of the reflected solar energy and F is a purely geometrical factor which is a function of the altitude of the body and its orientation relative to the planet-sun axis. F is plotted in Figure 12-25 as a function of altitude for various values of the bistatic angle σ. The bistatic angle is formed by the planet-sun line and the position vector of the satellite. F has been computed by Dennison for numerous orbit parameters.

12.4.3 Terrestrial Atmospheric Background

The radiance of the sky in the optical wavelengths is the result of two mechanisms: molecular scattering of incident radiation and emission by atmospheric constituents as a result of absorption of incident radiation. Atmospheric emission is significant only at wavelengths longer than 2 μ. Scattering of solar radiation is the overwhelming contribution to daytime sky radiance in the visible and near visible wavelengths. Daytime spectral sky radiance versus wavelength for the infrared region is plotted in Figure 12-26[31] and for the visible region in Figure 12-27. Figure 12-27 assumes clear sky conditions; radiances of sunlit clouds are typically an order of magnitude greater.

S.K. Mitra[32] estimates the contributions of various sources of night sky radiance as follows:

Starlight	30%
Zodiacal light	15%
Galactic light	5%
Airglow	40%
Scattered light from last 3 sources	10%

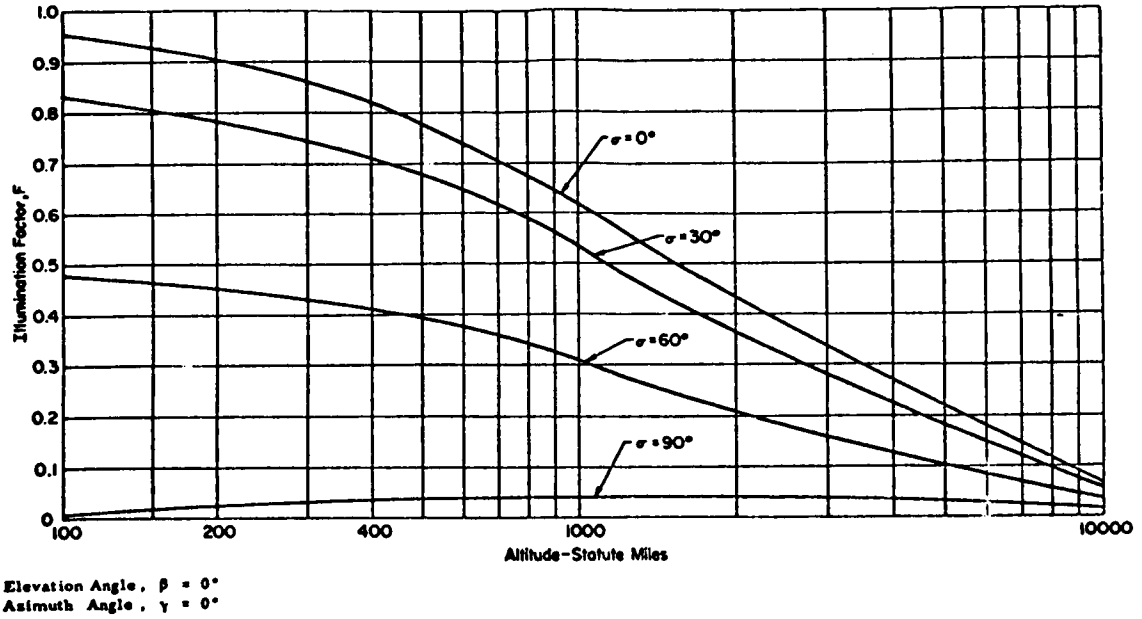


Figure 12-25. Illumination factor as a function of altitude for parametric values of bistatic angle, σ.

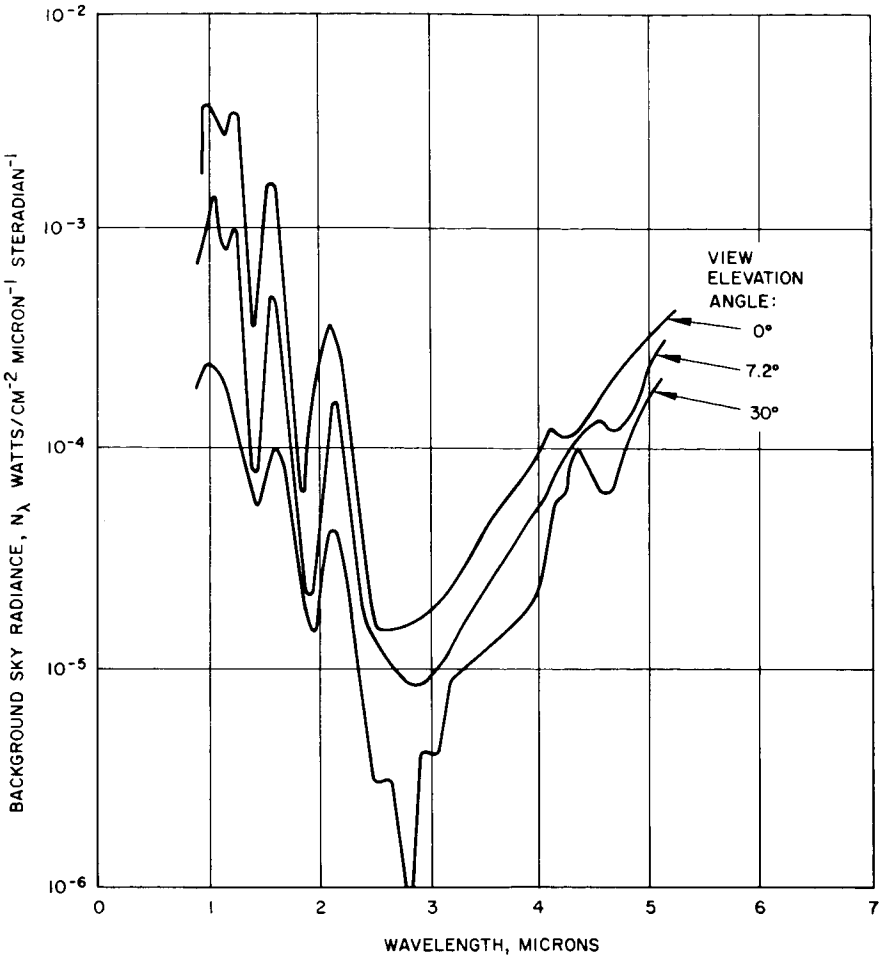


Figure 12-26. Clear sky radiance at Colorado Springs, Colorado; sun angle = 45 degrees.

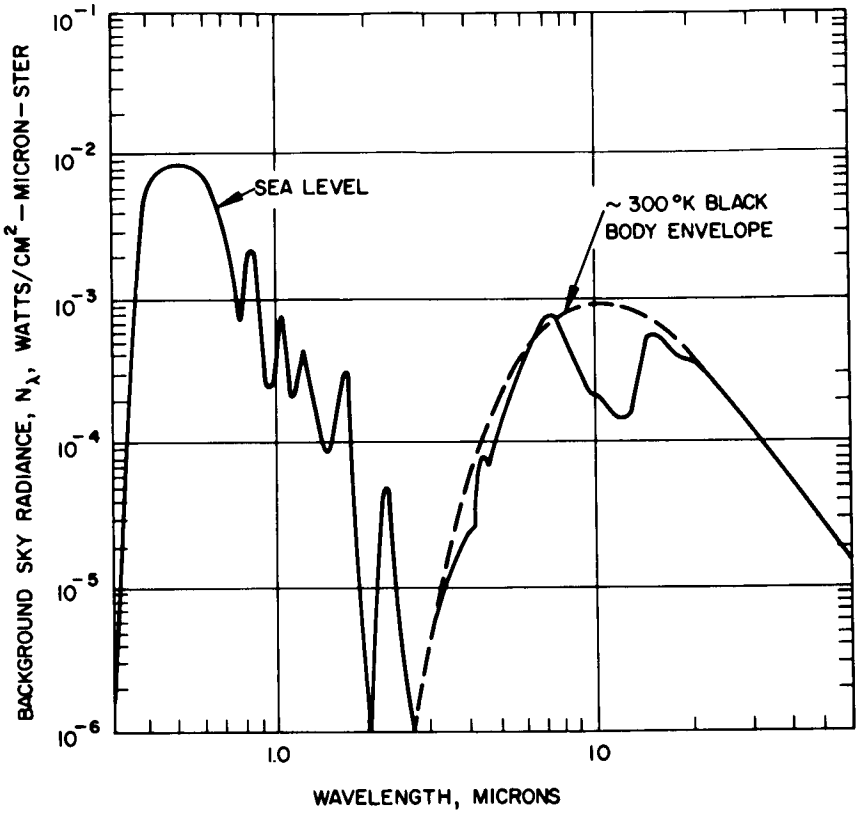


Figure 12-27. Diffuse component of typical background radiance from 30 km and sea level zenith angle 45 degrees, excellent visibility.

These estimates are for visible wavelengths and for conditions several hours after sunset with no moon. Twilight and moonlight will be considered separately. All these components of night emission vary in different ways with direction, time, atmospheric and meteorological conditions. Values of night sky radiance presented here are typical.

According to Mitra, peak sky radiance due to stellar sources is of the order of 10^{-1} watts/cm^2-steradian-μ at 0.55μ and follows approximately a Planckian spectral distribution. The effective irradiance produced at ground level is 3.34×10^{-10} watts/cm^2 at 0.55 . These figures are 1/4 to 1/6 the actual visible light observed from the night sky at a dark location on a clear night. The remaining contributions come from diffuse sources. The principal noise interference problem due to stars is the result of the relatively small number of very bright stars.

Measurements of the spectrum of diffuse night sky radiance have been made by F. E. Roach.[33] The results shown in Figure 12-28 are averaged over times, seasons, and direction of view in the celestial sphere. Figure 12-29[34] shows clear sky radiance in the infrared for various zenith angles. Except for the narrow intense N_a and H atomic lines, relatively light sky emissions appear between 0.1 and 1.0 microns. Beginning at 1.0 microns intense OH molecular bands appear as "air glow". Above 2 microns thermal emission from the dense lower atmosphere obscures the air glow. Twilight zenith sky radiance in the 0.4 to 0.7 micron band is illustrated in Figure 12-30[35] as an implicit function of the depression of the sun below the horizon. Zenith sky radiance in the 0.4 to 0.7 micron band from the full moon versus lunar elevation angle is illustrated in Figure 12-31[36]. Variation in lunar sky radiance at zenith with lunar phase is illustrated in Figure 12-32.[37]

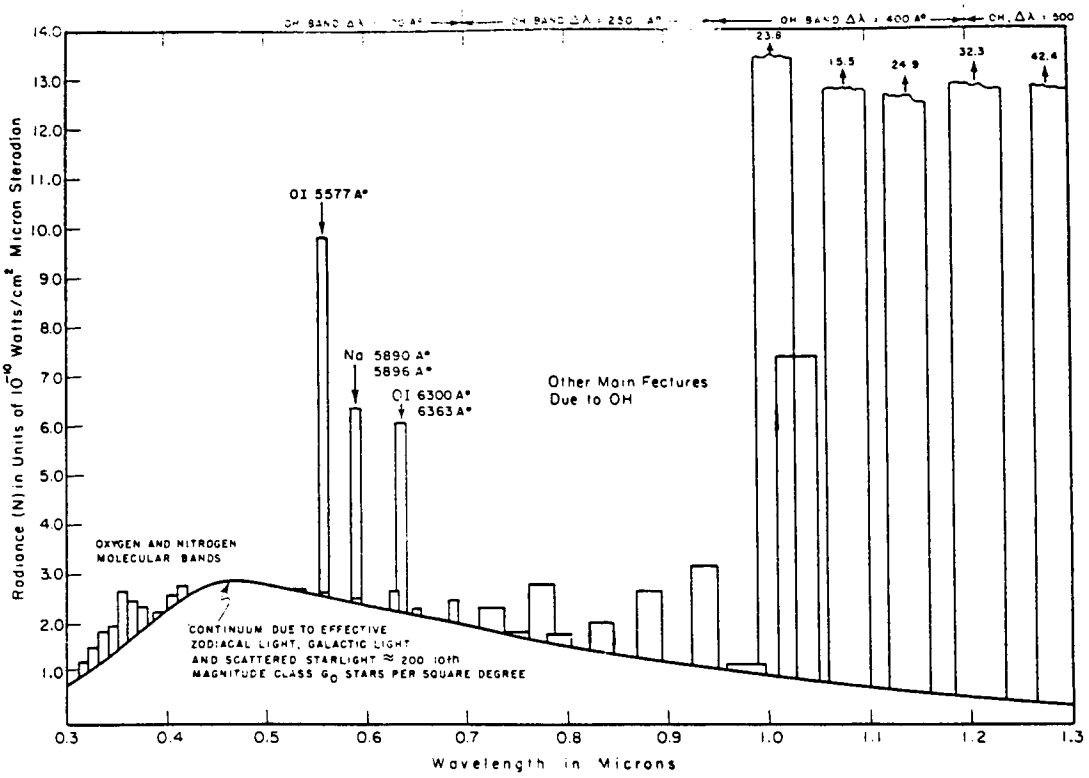


Figure 12-28. Estimated average diffuse radiation from zenith night sky. Average diffuse radiation from zenith, 0.3 to 1.3 microns.

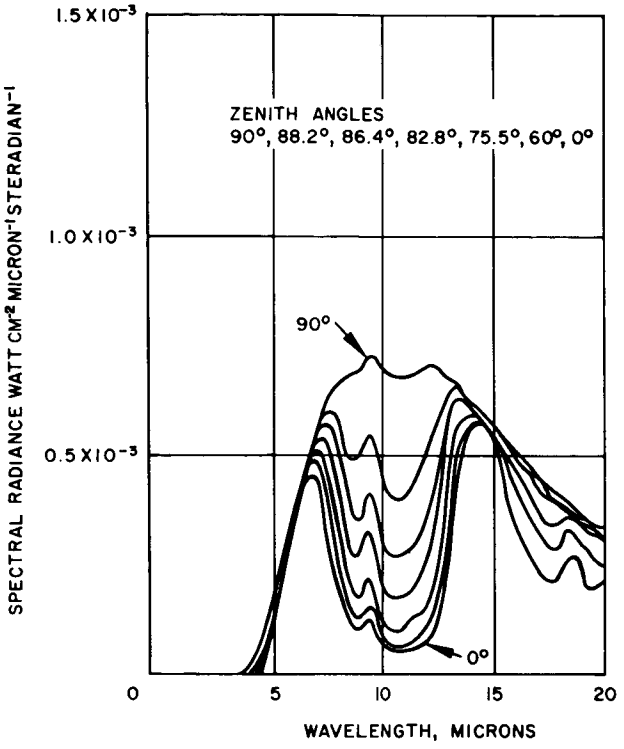


Figure 12-29. Spectral radiance in the infrared of a clear sky for several elevation angles above the horizon. These spectra were measured at an elevation of 11,750 feet, at night, with an ambient temperature of 8°C.

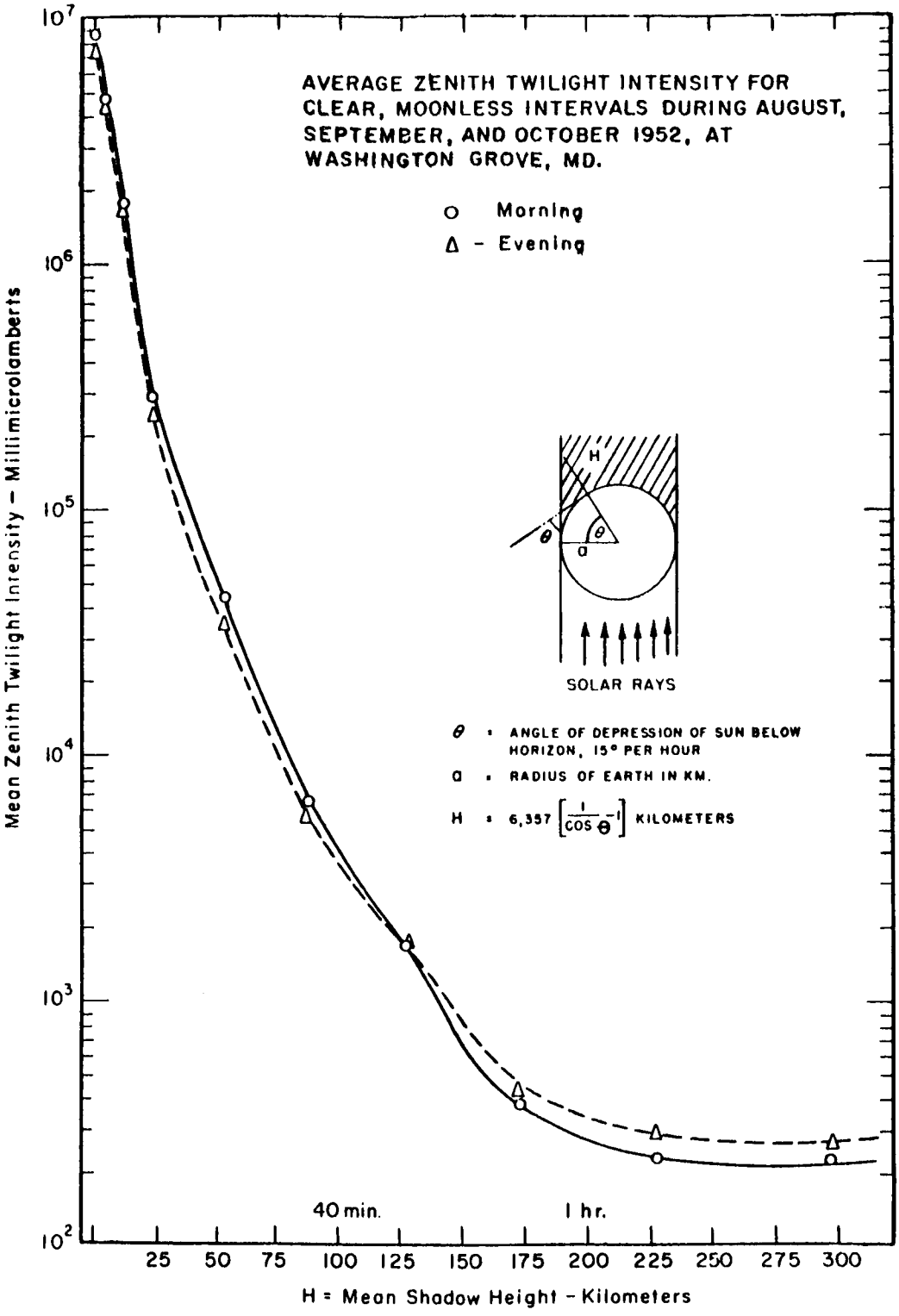

AVERAGE ZENITH TWILIGHT INTENSITY FOR CLEAR, MOONLESS INTERVALS DURING AUGUST, SEPTEMBER, AND OCTOBER 1952, AT WASHINGTON GROVE, MD.
O Morning
Δ - Evening
Mean Zenith Twilight Intensity – Millimicrolamberts
10^7
10^6
10^5
10^4
10^3
10^2
H
θ
a
SOLAR RAYS
θ = ANGLE OF DEPRESSION OF SUN BELOW HORIZON, 15° PER HOUR
a = RADIUS OF EARTH IN KM.
H = 6,357 $\left[\frac{1}{\cos\theta}-1\right]$ KILOMETERS
40 min.
1 hr.
25 50 75 100 125 150 175 200 225 250 275 300
H = Mean Shadow Height - Kilometers

Figure 12-30. Zenith twilight intensity.

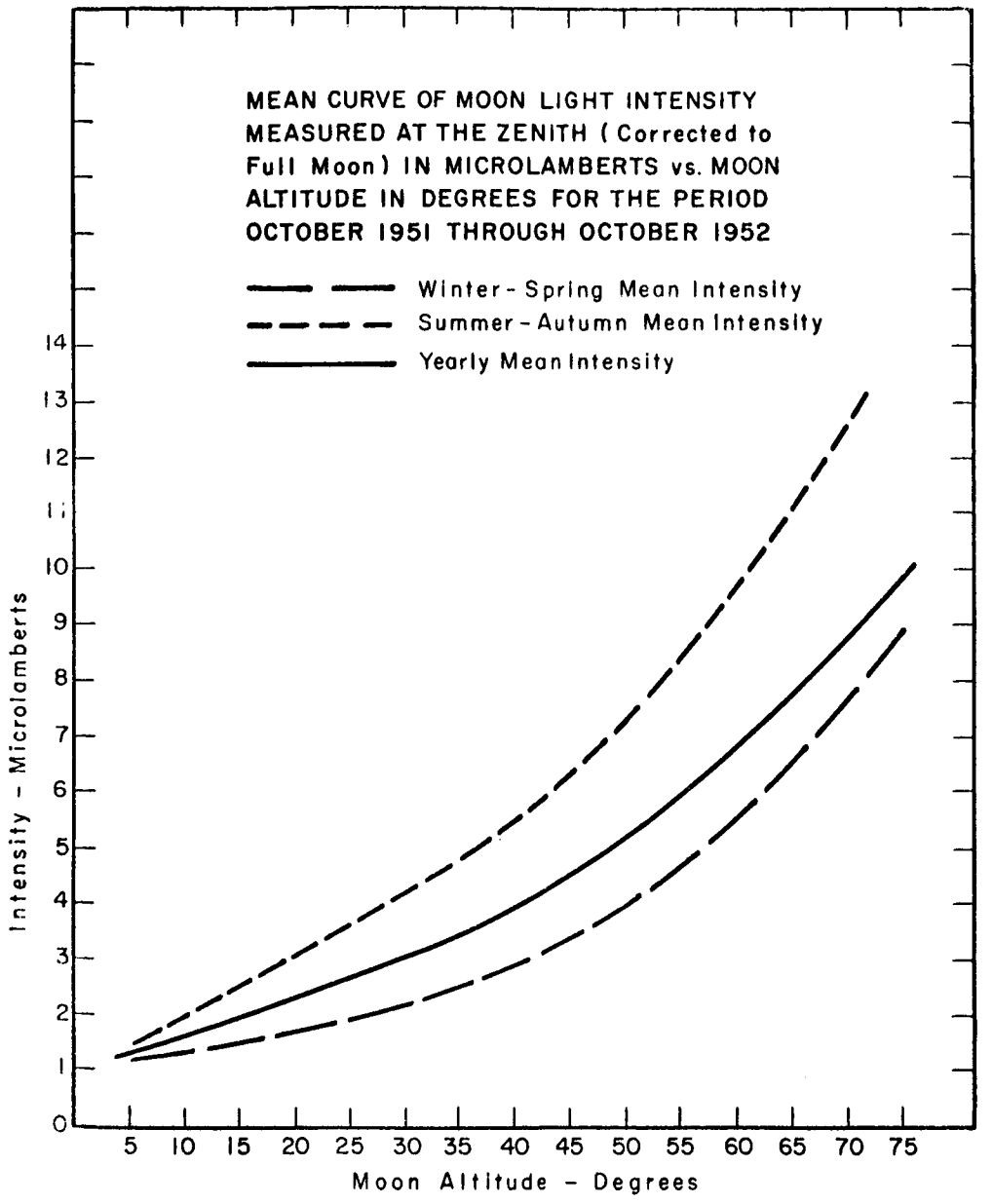


Figure 12-31. Zenith moonlight intensity.

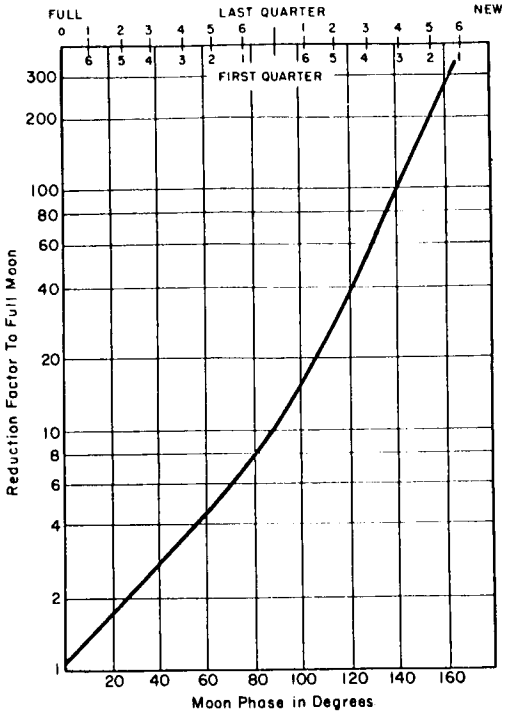


Figure 12-32. Correction factor for moon phase.

12.5 PROPAGATION OF RADIO FREQUENCY RADIATION THROUGH THE TERRESTRIAL ATMOSPHERE

12.5.1 Attenuation

Absorption in the ionosphere as illustrated in Figure 12-33[38] is very small for microwave frequencies. The effect of ionospheric attenuation can be approximated by assuming that each 0.1 db of attenuation is equivalent to 7°K antenna noise temperature. Ionospheric absorption is negligible for frequencies greater than 0.3 gcps.

Tropospheric absorption is due almost entirely to water vapor. Calculated attenuation curves for the frequency range 1.1 gcps to 100 gcps are shown in Figure 12-34[39] for various values of zenith angle. The curves are in satisfactory agreement with experimental attenuation data, shown in Figure 12-35[40]. A summary propagation data for a horizontal path, including rain and fog is given in Figure 12-36.

12.5.2 Turbulence Effects

For rf propagation through the troposphere, variations in the index of refraction are principally due to water vapor irregularities. These irregularities can be attributed to atmospheric turbulence. In the stratosphere, which contains a negligible amount of water vapor as compared to the troposphere, variations in the refractive index are less and can be assumed to be primarily a function of thermodynamic variations in pressure and temperature. Models have been developed, describing the atmospheric structure in statistical terms, as a basis for the application of scattering theories[41, 45]. Some correlation has been achieved between these theories and a number of experimental measurements[46, 51] taken over various path lengths under different and incompletely defined meteorological conditions. A more definitive experiment is required to provide good correlation between meteorological conditions and measured effects on propagation. Such a program is presently underway, conducted by the Stanford Research Institute, and is designed specifically to determine propagation limitations on multiple aperture antennas for deep space communications. However, at this time only the

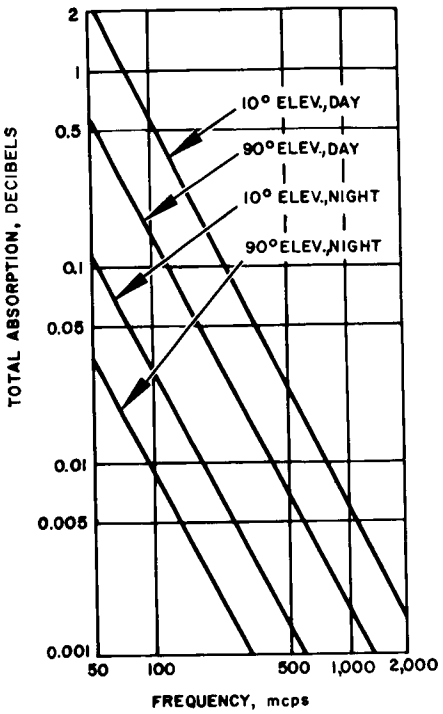


Figure 12-33. Ionospheric attenuation for a source at 1,000-km height.

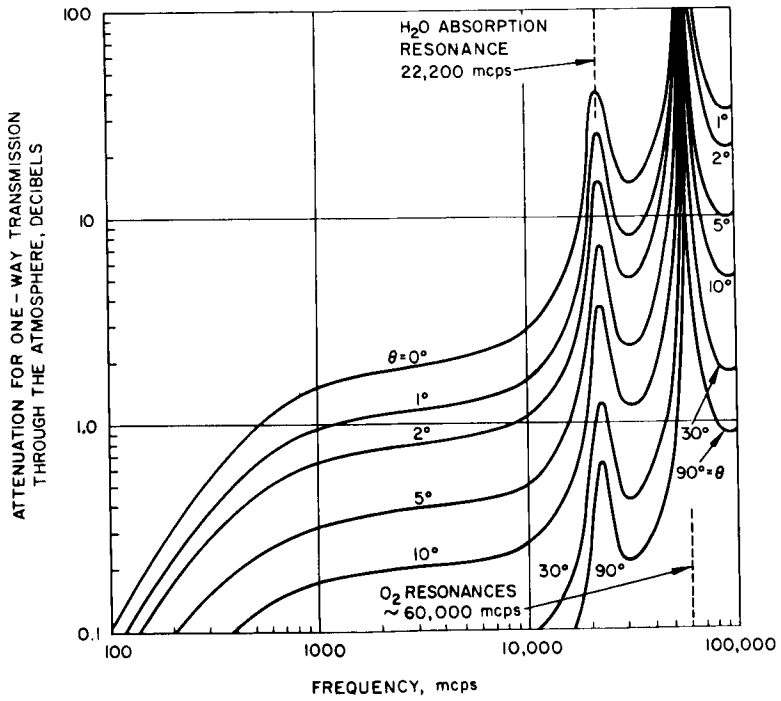


Figure 12-34. One-way attenuation through the standard summer atmosphere due to oxygen and water vapor.

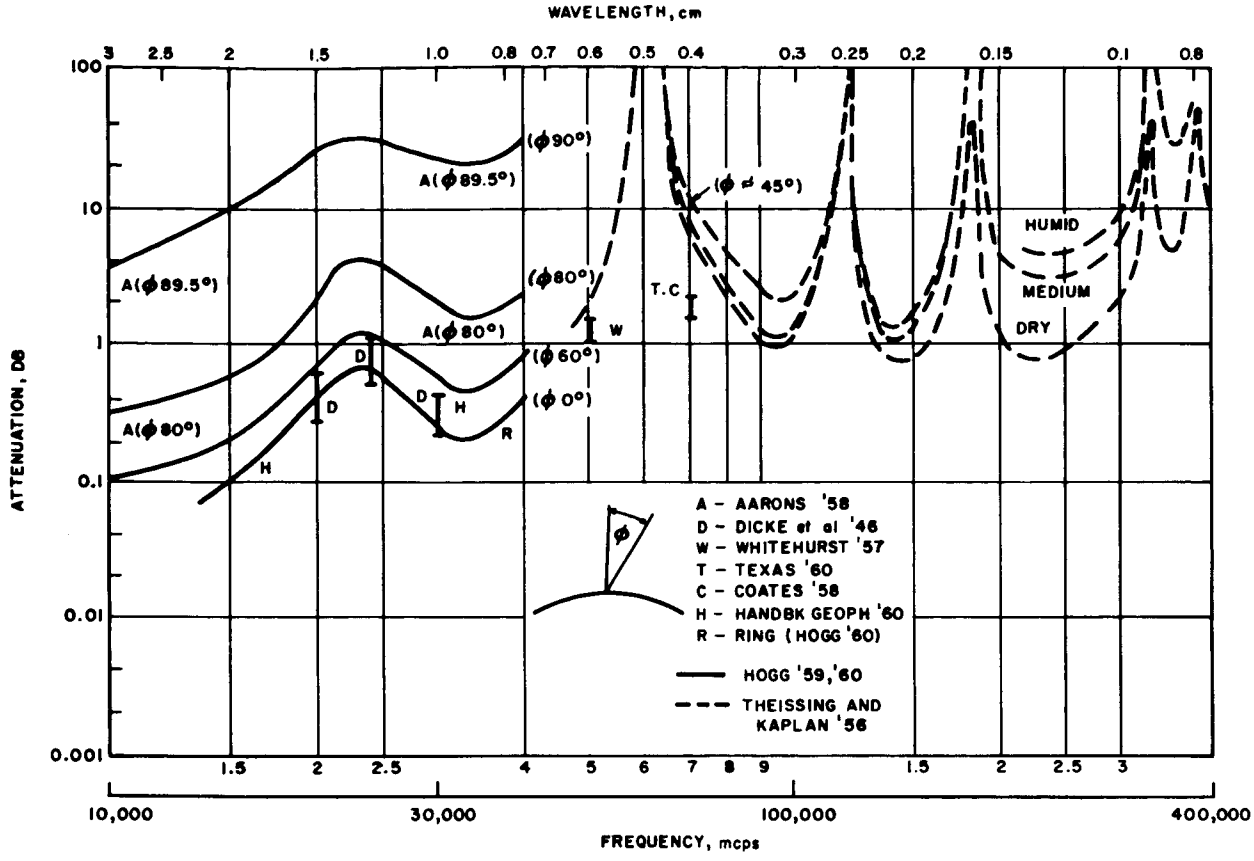


Figure 12-35. Total attenauation for one-way transmission through the atmosphere. (Proc. IEEE, P. 488, April 1966)

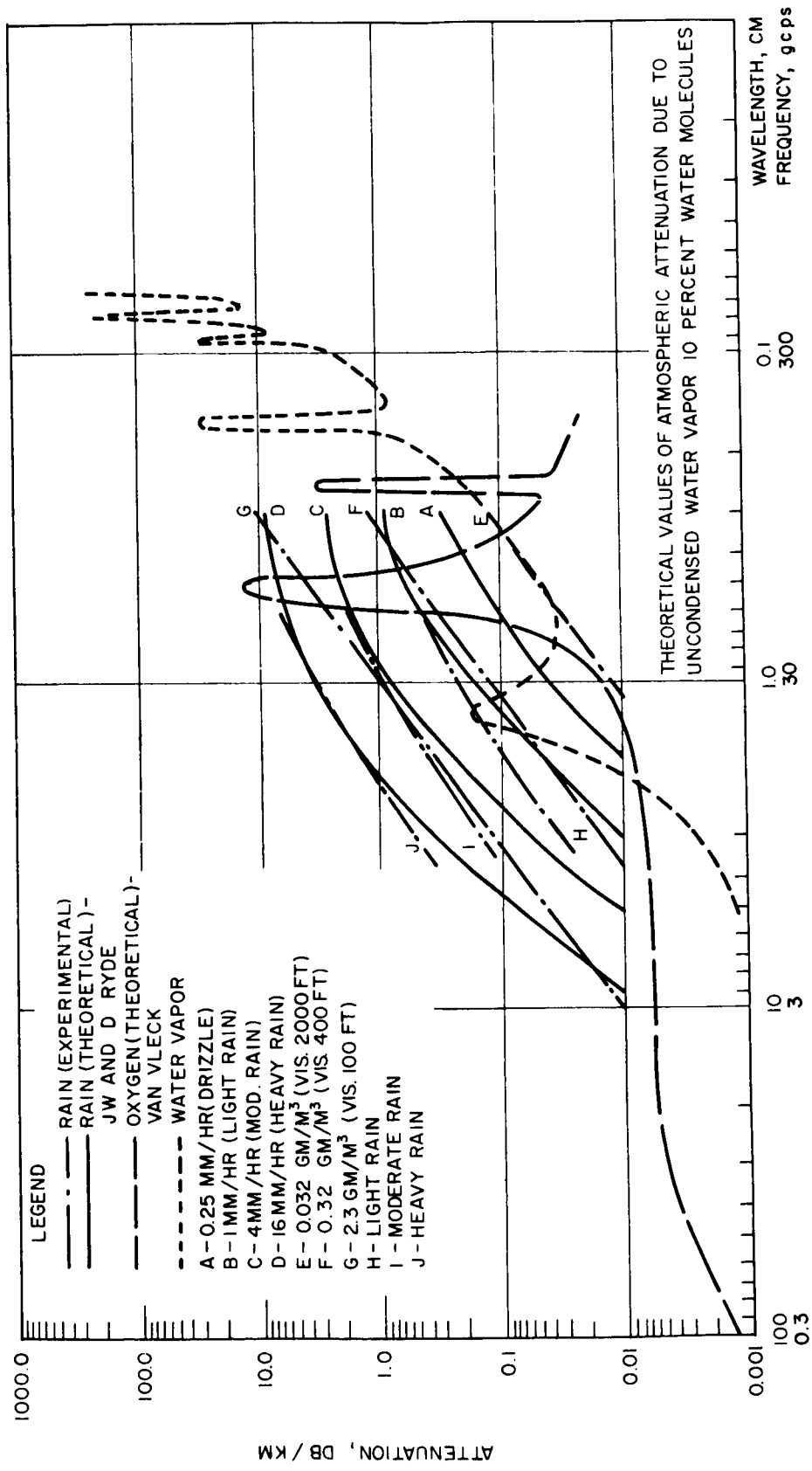


Figure 12-36. Atmospheric attenuation summary.

preliminary analyses have been conducted[52] to outline the extensive experimental program; the tasks of equipment construction, data collection, and evaluation remain.

In the absence of definitive data, an indication of expected deviations of an incoming wavefront over an antenna aperture can be gained from the phase measurements obtained in the NBS Maui experiment[49]. Here phase deviations in transmission over a 15-mile path, which dropped from 10,000 feet to 100 feet, were measured by two receivers for several baseline lengths up to 4,800 feet. The measurements were made at a frequency of 9,414 mcps, but the deviations in terms of path length can be expected to be relatively independent of frequency. Since the measurements at the various antenna locations were made during different recording periods and therefore under different meteorological conditions, the sample points are not strictly correlated over the baseline range. The trend of these samples, however, is remarkable consistent and is shown in Figure 12-37 where phase deviations have been converted to linear deviations in the wavefront.

12.6 PROPAGATION OF OPTICAL FREQUENCY RADIATION IN THE TERRESTRIAL ATMOSPHERE

12.6.1 Attenuation

Attenuation by the atmosphere at optical wavelengths is due to absorption by molecular atmospheric constituents and to molecular and particle scattering. The entire wavelength region from 0.3 to 5μ contains thousands of sharp absorption lines due to H_2O, CO_2, N_2O, CH_2, O_2, O_3, CO and their isotopes. At low resolution these lines are smoothed out so that they are manifested as absorption bands. Scattering attenuation due to very small particle or molecular scattering is called Rayleigh scattering. Rayleigh scattering varies as $1/\lambda^4$ and produces large attenuation at short wavelengths. Scattering by particles large relative to the wavelength is called Mie scattering and is much less wavelength dependent. Figures 12-38 and 12-39[53] present a low resolution picture of atmospheric attenuation from 0.3 to 5μ. These

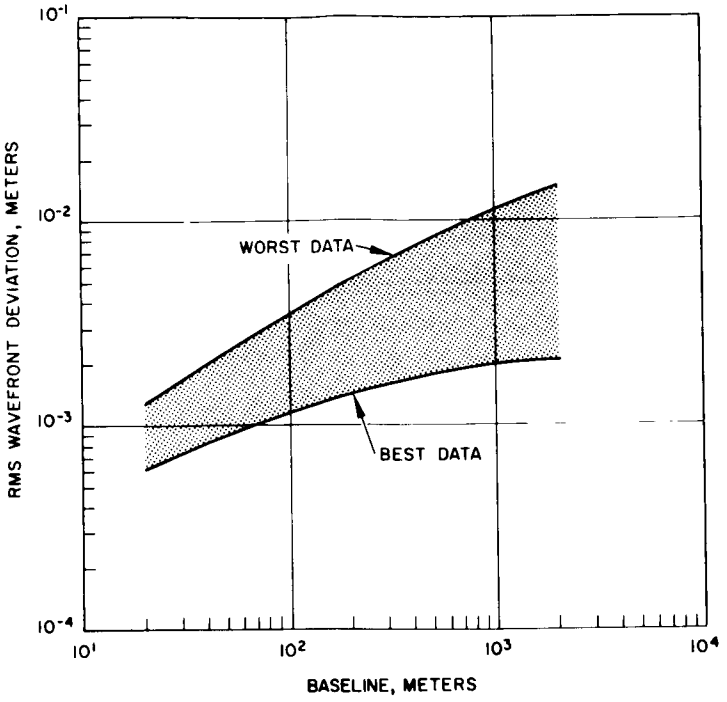


Figure 12-37. Atmospheric distortion of r-f wavefront based on NBS main data.

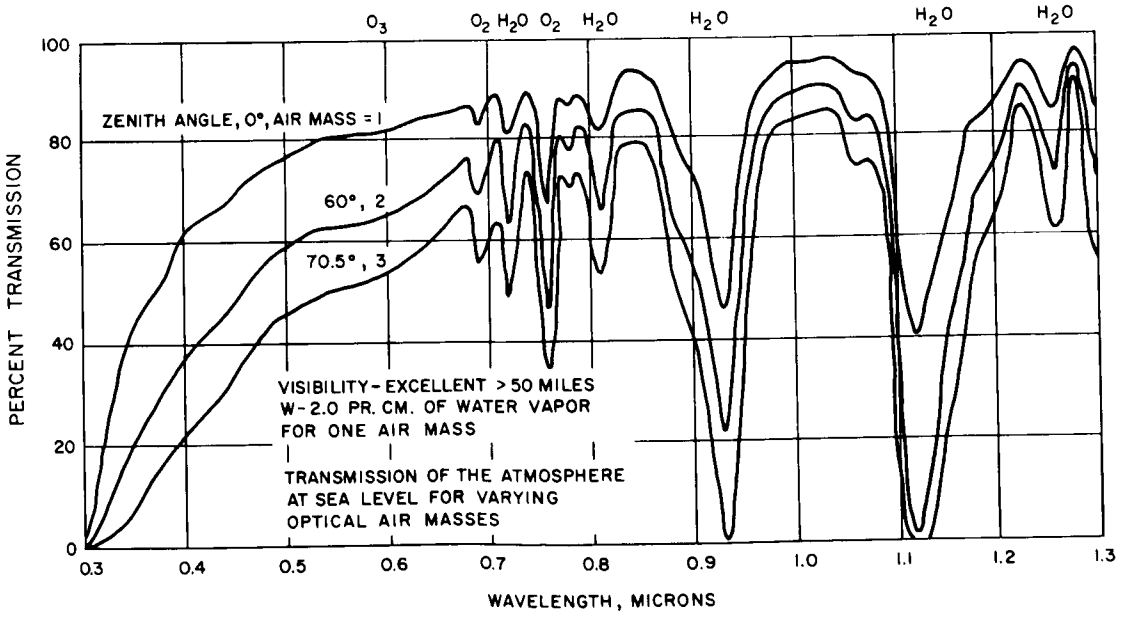


Figure 12-38. Atmospheric transmission, 0.3 to 1.3 microns.

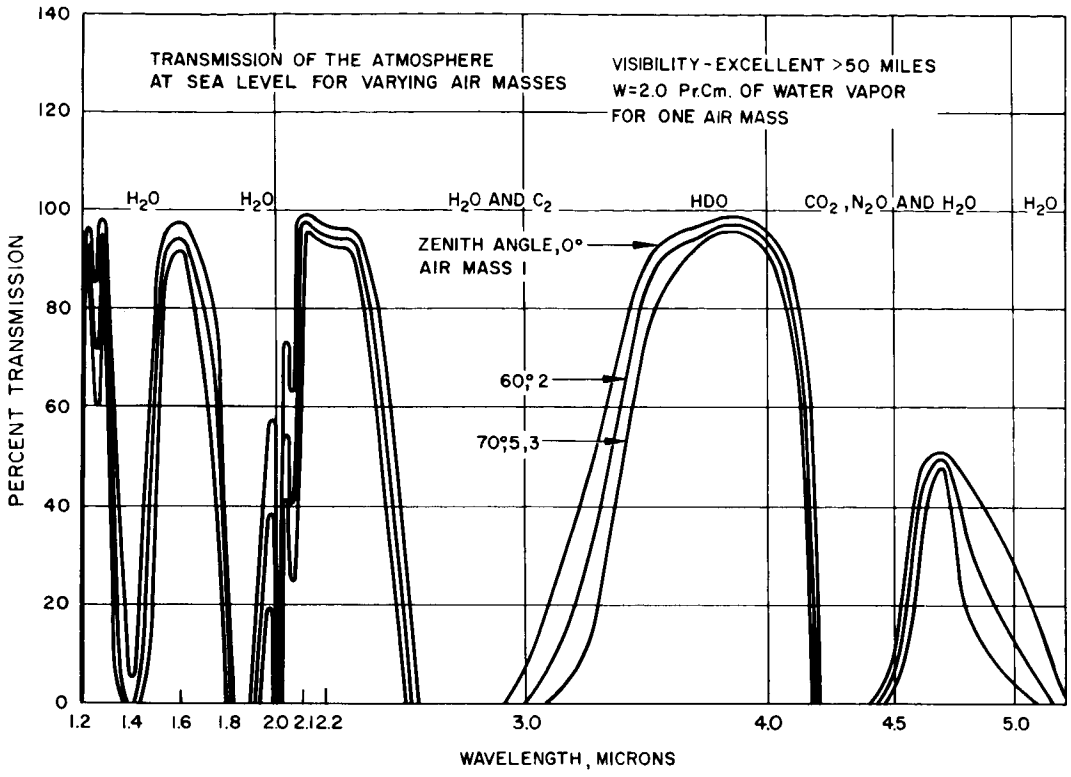


Figure 12-39. Atmospheric transmission, 1.2 to 5.0 microns.

curves are based upon excellent sea level visibility (> 50 miles) and typical atmospheric water vapor content (2 precipitable centimeters of water in a vertical column above the observation station). Attenuation at 0.63μ through fog and rain is shown in Figure 12-40[54].

12.6.2 Turbulence Effects

Communication and tracking by coherent light through the atmosphere results in system limitations in addition to those imposed by attenuation and scattering associated with incoherent waves. Laser beam directivity and coherence are degraded by the presence of atmospheric turbulence. The turbulent atmosphere is inhomogeneous and its index of refraction is a function of position and time. As a result of this variation, the character of the beam is altered by the following phenomena:

1. Loss of spatial coherence: destruction of the phase coherence across the beam, the phase changing rapidly with position, leads to blurring of the image.
2. Scintillation: Variations in intensity due to atmospheric turbulence result from random changes in beam cross section.
3. Image dancing: variations in the angle of arrival of the received wavefront will cause the image to be focused at different points in the focal plane of the receiving optics.
4. Beam steering: the entire beam may be deviated from the line of sight, giving rise to a loss of power at the receiving aperture.
5. Random polarization fluctuation.

Any quantitative treatment of the effect of atmospheric turbulence must account for the finite beam diameter, D, which may vary from several millimeters up to several meters, and the important parameter is the ratio of D to inhomogeneity dimensions ℓ.

First, if $D/\ell \ll 1$, the major effect of the turbulence will be to refract the beam as a whole, so that over ranges large compared to ℓ, the projected location of the center of the beam will execute a two-dimensional random walk in the receiver plane. Hence, if the turbulence is isotropic, the displacement of the beam from the line of sight will be

Rayleigh distributed. Second, if $D/\ell \approx 1$, the inhomogeneities will, in a first approximation, act as lenses which focus and defocus all or parts of the beam, causing a granular structure of the wavefronts. The steering and spreading of the beam will both be small in this case; and unless the transmission range is larger than the far-field range, scintillation will also be small. Last, if $D/\ell >> 1$, small portions of the beam will be independently diffracted and scattered (focused and defocused). Thus, for ranges large compared to ℓ^2/λ, along with spreading of the beam, we expect a badly distorted wavefront in which intensity fluctuations are strong.[55] For an earth-based receiver and a deep space transmitter, $D/\ell >> 1$ is always the case so that the principal effects are loss of coherence, scintillation and image dancing. For an earth based transmitter and spaceborne receiver, $D/\ell << 1$ is the case and image dancing and scintillation result.

12.6.3 Loss of Spatial Coherence[56]

Spatial coherence is degraded by phase changes in a direction perpendicular to the propagation direction. Due to local refractive index variations in the path of the rays collected by the aperture the initially plane wavefront becomes distorted so that the aperture is no longer an equiphase surface.

In order to illustrate the effects of loss of spatial coherence at the aperture plane of an optical heterodyne receiver, a power reduction factor ρ will be defined as the ratio of the power received when the wave is partially coherent to that received when it is totally coherent. ρ may be expressed for optical heterodyne detection systems as

$$\rho = \frac{1}{z^2}\psi(z) \qquad (12\text{-}32)$$

where $\psi(z)$, plotted in Figure 12-41 (Fried) contains the dependence of signal-to-noise ratio on receiver diameter and z is the receiver aperture diameter in units of coherence length r_o, given approximately by

$$r_o(\lambda, \phi) = r_o(\lambda_o, 0)\left(\frac{\lambda}{\lambda_o}\right)^{6/5}[\cos \phi]^{3/5} \qquad (12\text{-}33)$$

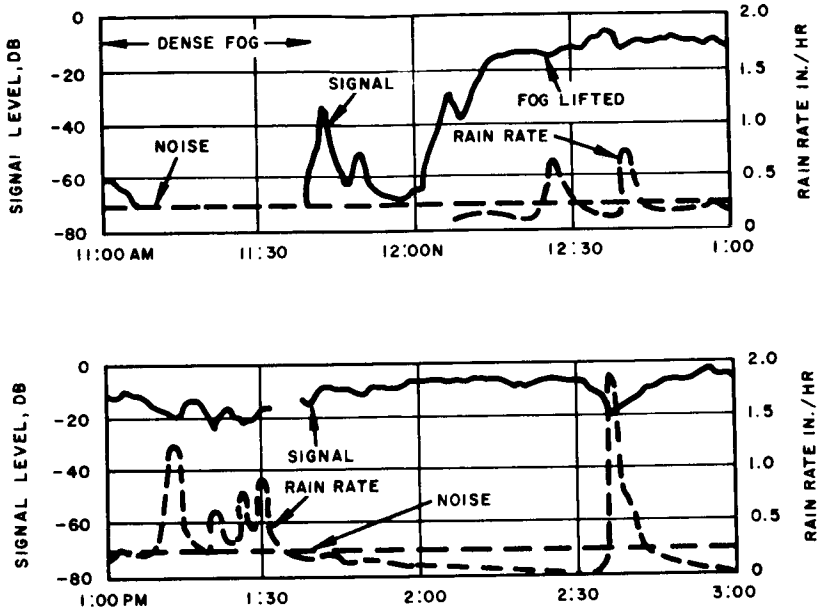


Figure 12-40. Fog-rain attenuation, $\lambda = 0.63\mu$.

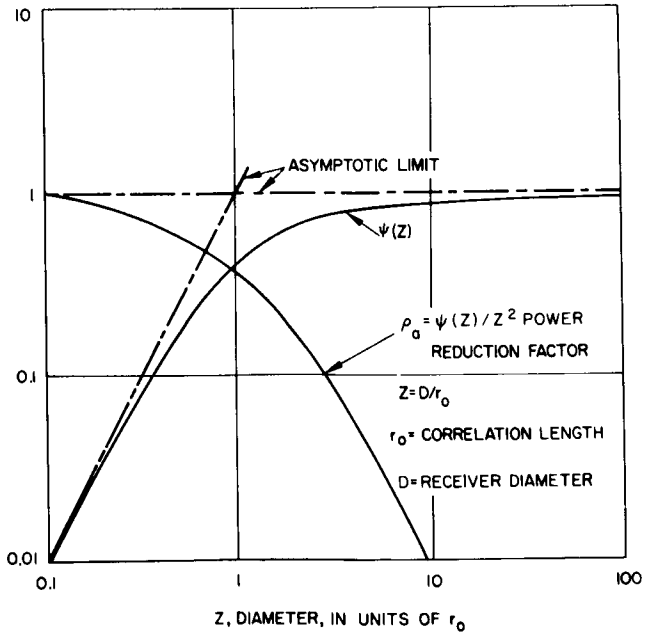


Figure 12-41. Power reduction caused in an optical hetrodyne receiver caused by loss of phase coherence (after Fried).

where

λ = wavelength

ϕ = zenith angle

$r_o(x_o, 0)$ = coherence length for wavelength λ_o and zenith angle 0.

Physically, r_o is a measure of the spatial coherence in a plane perpendicular to the propagation direction. The power reduction factor ρ decreases as z^2 for $z > 1$. The power reduction factor is plotted in Figure 12-41 as a function of normalized receiver aperture diameter. The coherence length r_o is plotted in Figure 12-42 as a function of wavelength for various receiver altitudes and transmission path zenith angles (as indicated geometrically in Figure 12-43).

It can be seen that the useful collector diameter may be quite small for wavelengths in the visible and near infrared. The advantage of using longer wavelengths can be seen from the $\lambda^{6/5}$ dependence of the coherence length. Since this is approximately a linear relationship, a factor of 10 or more increase in effective collecting aperture diameter is possible in going from 1.06μ to 10.6μ.

12.6.4 Image Dancing

For the case of an earth receiver viewing an exoatmospheric plane wave source, Hufnagel[57] has shown that the mean square beam deflection on the image plane of a receiver having aperture D (cm) is given by

$$\overline{\Delta\theta^2} = \frac{2.7 \times 10^{-10} \sec\phi}{D^{1/3}} \qquad (12\text{-}34)$$

where ϕ is the zenith angle. Table 12-3 gives bounds on the rms beam deflection for average daytime seeing conditions for various zenith angles. Diffraction limited optics and a source wavelength of 6328Å are assumed.

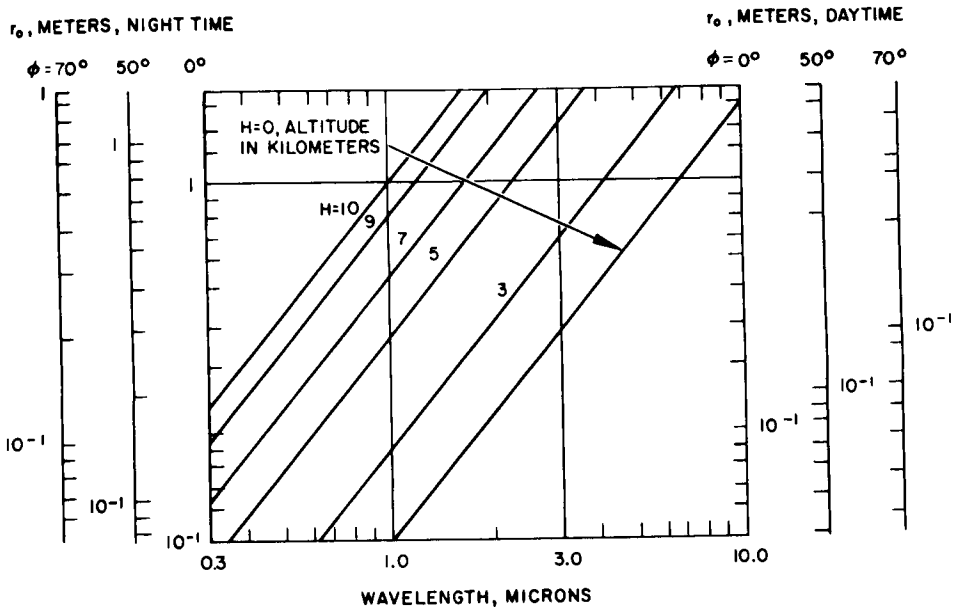


Figure 12-42. Dependence of coherence length, γ_o on zenith angle, ϕ, and wavelength, λ_μ (after Fried).

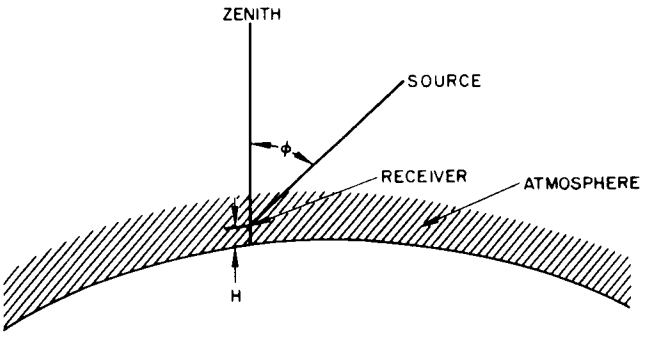


Figure 12-43. Geometry for optical heterodyne receiver in the atmosphere (after Fried).

RMS angle of arrival fluctuations versus zenith angle ϕ and receiving aperture diameter are listed below from Equation (12-34)

Diffraction limited beamwidth	1 sec	0.1 sec
Receiving aperture diameter	0.131 m	1.31 m

Zenith Angle	$\Delta\theta_{rms}$ arc-sec (μ rad)	
0	1.75 (8.4)	0.76 (3.65)
30	1.88 (9.0)	0.82 (3.84)
45	2.08 (10.0)	0.91 (4.37)
60	2.45 (11.8)	1.08 (5.18)

Table 12-3. Beam deflection limits for average daytime seeing conditions at various zenith angles.

The random variation in angle of arrival of the beam in the earth based is analogous to the "quivering" of stellar images. It is found that stellar images do not remain fixed in the telescope field of view, but undergo random displacements in all directions. Typical observed values of $\Delta\theta_{rms}$ as a function of zenith angle are shown in Figure 12-44[58] for nighttime turbulence conditions.

Davis[59] states that for typical daytime turbulence conditions, the rms deviation in angle of arrival $\Delta\theta_{rms}$ for vertical propagation through the atmosphere is approximately 14 μ rad. Weak and strong turbulence conditions change this by factors of 0.2 and 1.25 respectively. For nighttime turbulence conditions $\Delta\theta_{rms}$ is stated to be approximately 2.1 μ rad. As was pointed out, the effect of imaging dancing is to cause an image to be focused at different points on the focal plane of the receiving optics. However, for the case of an earth based transmitter where $D/\ell \ll 1$, it leads to beam steering and scintillation in the far field (i.e., at the deep space receiver).

12.6.5 Scintillation

Scintillation or fluctuations in intensity of signals received from a far exoatmospheric optical transmitter are analogous to the twinkling of

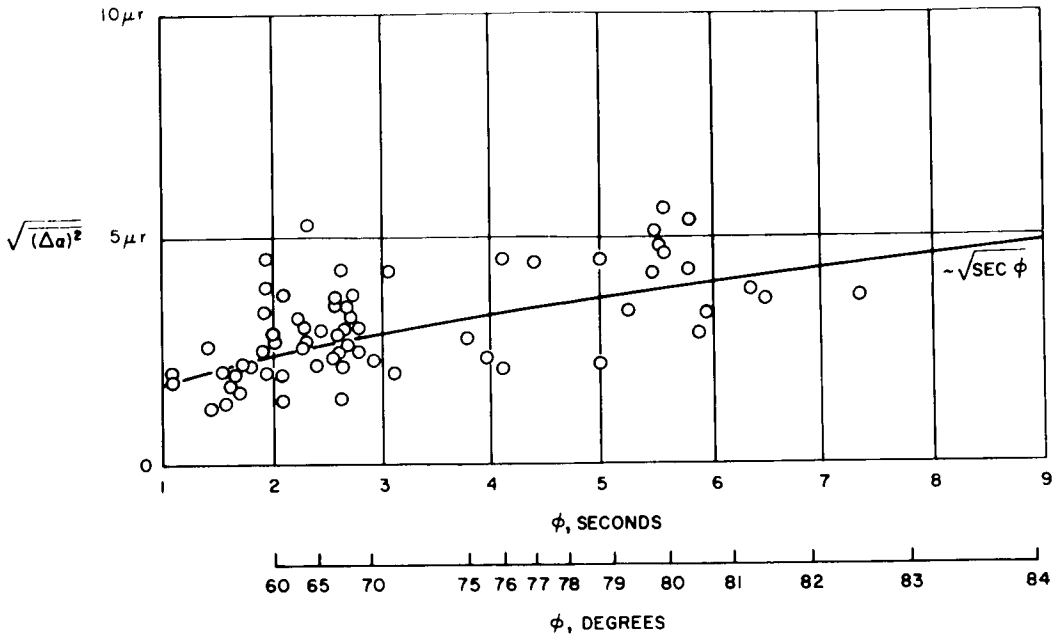


Figure 12-44. Dependence of the amount of quivering of stellar images on the zenith distance.

stars for which considerable empirical data at visible wavelenghts is available. The magnitude of the fluctuations depends on the dimensions of the receiving optics aperture, the zenith angle of the light source, and meteorological conditions. The receiving aperture diameter also significantly effects the dependence of fluctuation magnitude on source zenith angle. Figure 12-45 shows the empirical dependence of scintillation amplitude on receiving aperture diameter. Figures 12-46 and 12-47 show for two receiver aperture sizes variation of the quantity log σ_p with zenith angle ϕ. Here $\sigma_p = [(P - \overline{P})/\overline{P}^2]^2$, P is the actual light flux through the receiver aperture, and $\overline{P}$ is the time average light flux through the aperture.[60]

For propagation from a receiver located at an altitude h through the entire atmospheric layer, Davis[61] has developed graphs showing scintillation as a function of wavelength. Figures 12-48 through 12-50 show for average daytime turbulence conditions the mean square deviation in log level of intensity versus wavelength λ for zenith angles of 0, 60, and 70.5 degrees respectively. These curves are valid for $L_o \gg \sqrt{\lambda R}$ where L_o is the correlation length or outer scale size for atmospheric turbulence, λ is wavelength, and R is range.

12.6.6 Beam Steering

The net effect of numerous deflections of the beam over its path is to produce a deviation $\Delta\alpha$ of the beam from its line of sight path. This effect is referred to as beam steering. Essentially the projected location of the center of the beam executes a two-dimensional random walk in the receiver plane. If the turbulence is isotropic, the displacement of the beam from the line of sight will be Rayleigh distributed. Making the assumption that the correlation distance for the derivative of the angular deviation with respect to distance along the path is small compared to the total range R, it can be shown that the net lateral displacement of a ray has an rms deviation

$$R\Delta\alpha_{rms} = \frac{R\Delta\theta_{rms}}{\sqrt{2}} \qquad (12\text{-}35)$$

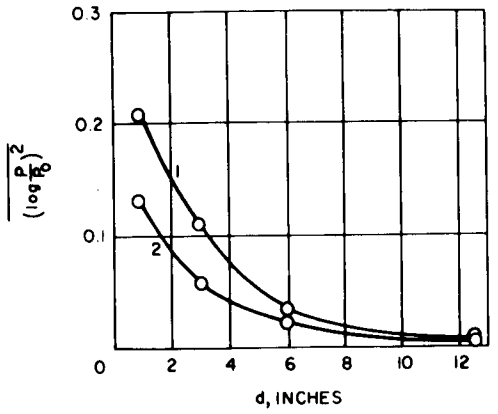


Figure 12-45. Empirical dependence of the amount of star twinkling on the diameter of the telescope diaphragm (1, winter; 2, summer).

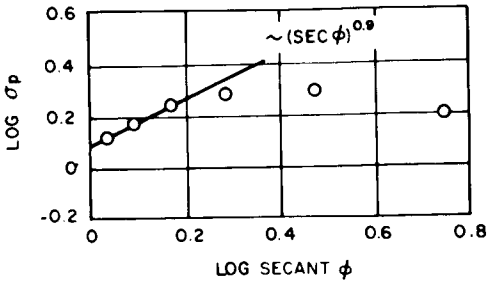


Figure 12-46. Dependence of the amount of star twinkling on the zenith distance when the telescope diaphragm has a diameter of three inches.

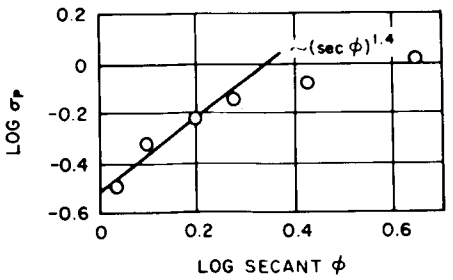


Figure 12-47. Dependence of the amount of star twinkling on the zenith distance when the telescope diaphragm has a diameter of 12.5 inches.

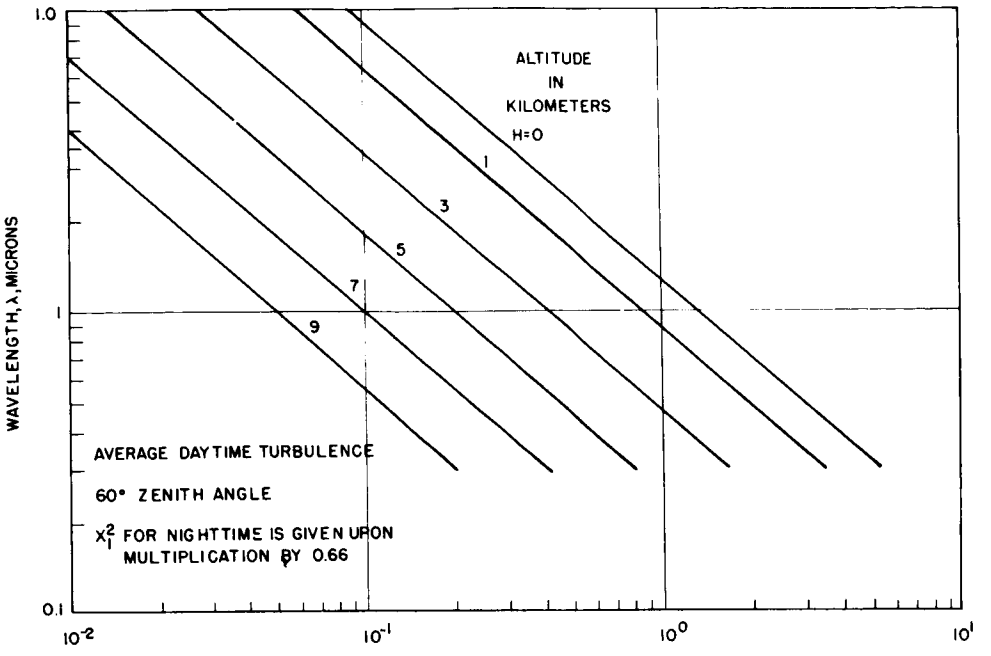


Figure 12-48. Mean-square deviation in log-level of intensity, X_I^2.

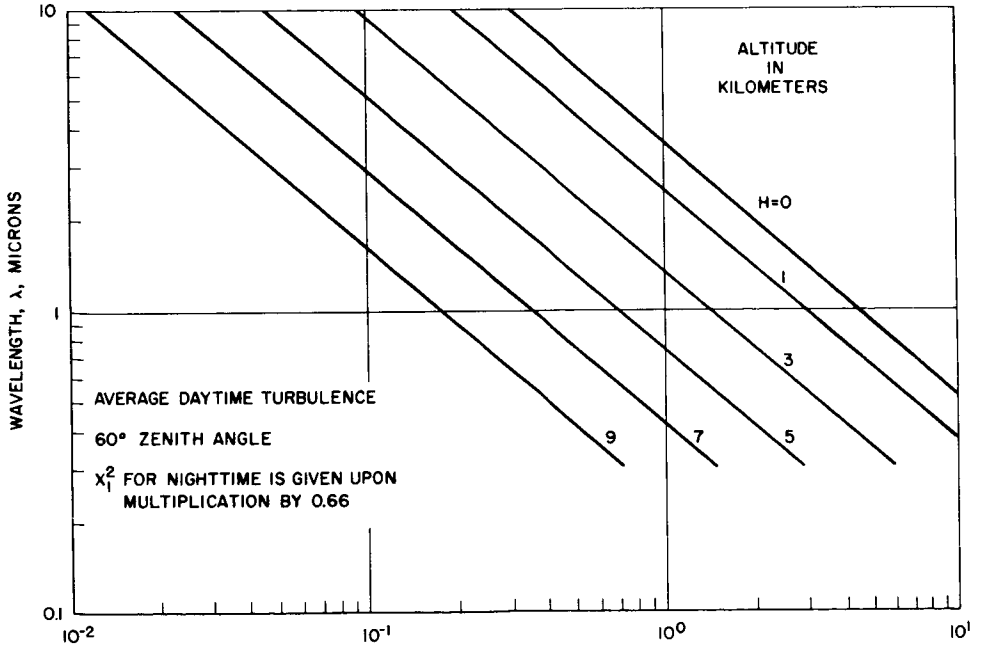


Figure 12-49. Mean-square deviation in log-level of intensity, X_I^2.

Hence it may be inferred that the beam is deviated by[62]

$$\Delta\alpha_{rms} = \frac{\Delta\theta_{rms}}{\sqrt{2}} \quad . \tag{12-36}$$

When the receiver is in the far field of the transmitter, beam steering will be negligible if the rms angular deviation of the beam $\Delta\alpha_{rms}$ is small (1/3 or less) compared to the half angular divergence of the beam.[56]

12.6.7 Random Polarization Fluctuation

An estimate the amount of polarization fluctuation to be expected has been made by Hodara.[63] If the initial polarization has components E_{1x} in the plane of incidence with the interface of constant index of refraction (See Figure 12-51) and E_{1y} parallel to that interface, the initial polarization direction is given by the argument of

$$\tan\phi = \frac{E_{1x}}{E_{1y}} \tag{12-37}$$

After refraction through the interface, the transmitted components have different polarization direction given by

$$\tan(\phi - \Delta\phi) = \frac{E_{2y}}{E_{2x}} \tag{12-38}$$

Let the correlation distance L_c be defined as the distance in which the spatial correlation function $C_{12}(\rho)$ of the atmospheric refractive index fluctuation Δn falls to 0.5.

$$C_{12}(\rho) = \langle \Delta n(r)\, \Delta n(r+\rho) \rangle \tag{12-39}$$

where the angular bracket indicates an ensemble average. Both L_c and $\langle \Delta n^2 \rangle$ are functions of altitude. Empirically it has been found

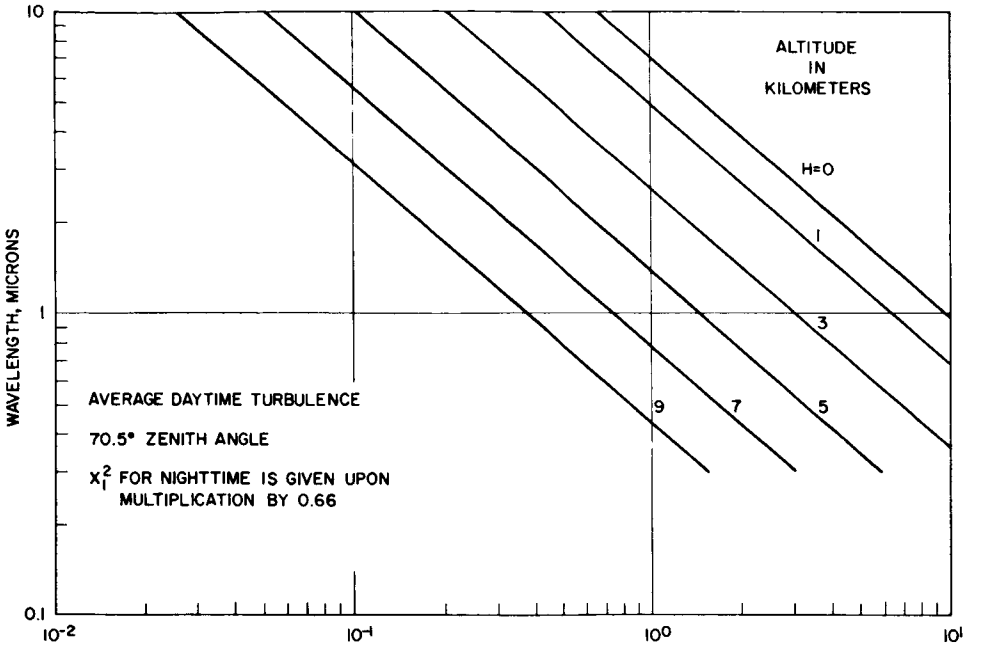


Figure 12-50. Mean-square deviation in log-level of intensity, X_I^2.

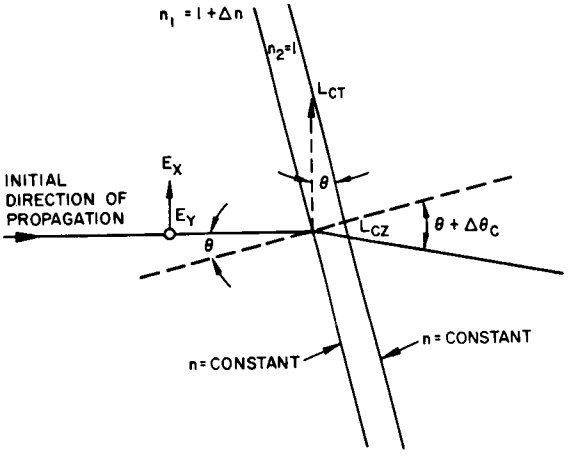


Figure 12-51. Geometry of propagation through a turbulent blob.

that $<\Delta n^2>$ and L_c are given by the curves of Figures 12-52 and 12-53. If L_{cz} is the correlation distance in the direction of propagation and L_{ct} is the correlation distance transverse to the direction of propagation, then for

$$\sqrt{(\Delta n^2)}\frac{L_{cz}^2}{L_{ct}^2} \ll 1$$

and (12-40)

$$\Delta n \tan^2 \theta \ll 1$$

the polarization components may be expressed in terms of the turbulence fluctuation parameters as

$$\frac{E_{2x}}{E_{1x}} = \frac{1 + \Delta n \quad (1 - \tan^2\theta)}{(1 + \Delta n/2)(1 - \tan^2\theta)} = x \qquad (12\text{-}41)$$

$$\frac{E_{2y}}{E_{1y}} = \frac{1 + \Delta n}{(1 + \Delta n/2)}(1 - \tan^2\theta) = y \qquad (12\text{-}42)$$

where θ is the angle of incidence at the interface of constant index of refraction. The change in polarization through the interface of constant index of refraction is

$$\Delta\phi_c = \tan^{-1}\frac{E_{2y}}{E_{2x}} - \tan^{-1}\frac{E_{1y}}{E_{1x}} \qquad (12\text{-}43)$$

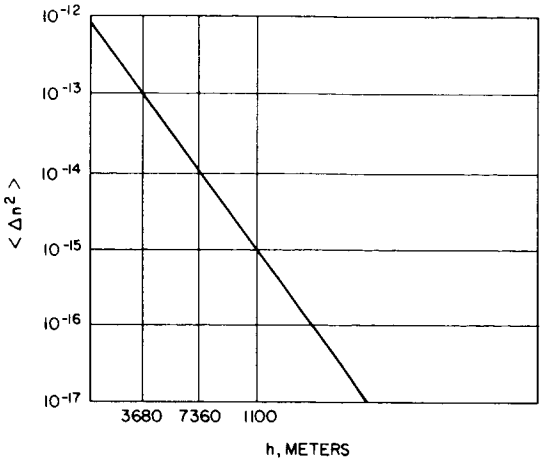


Figure 12-52. Mean square variation (Δn^2) vs. altitude, h.

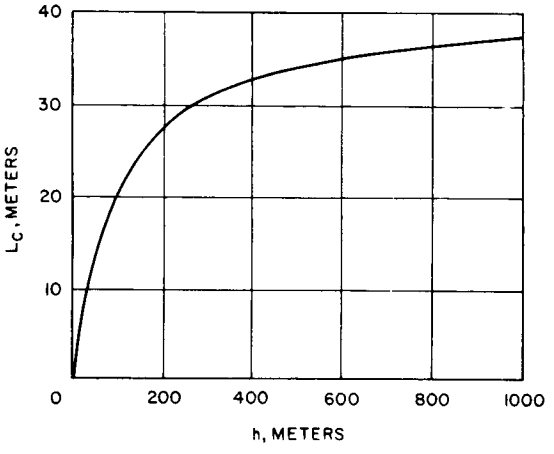


Figure 12-53. Correlation distance, L_o, vs. altitude, h.

In terms of turbulence fluctuation parameters

$$\Delta\phi_c = \frac{\left(1 - \frac{x}{y}\right) \tan \phi}{\tan^2 \phi + \frac{x}{y}} \qquad (12\text{-}44)$$

From Figure 12-51 it may be seen that $\tan \theta = L_{cz}/L_{ct}$

$$\tan \theta = L_{cz}/L_{ct} \qquad (12\text{-}45)$$

while from Equations (12-29) and (12-30)

$$\frac{x}{y} = 1 - \Delta n \tan^2 \theta \qquad (12\text{-}46)$$

Combining (12-40), (12-45), and (12-46) in (12-44) gives for the phase change in one coherence interval

$$\Delta\phi_c \sim \frac{\Delta n}{2} \frac{L_{cz}^2}{L_{ct}^2} \sin 2\phi \qquad (12\text{-}47)$$

Over a path length L there are L/L_c such elements. Adding the polarization changes on a mean square basis and taking the time average,* the mean square value of the maximum polarization change is given by

$$\Delta\phi = \sqrt{(\Delta n^2)} \frac{L_{cz}^2}{L_{ct}^2} \sqrt{\frac{L}{L_{cz}}} \qquad (12\text{-}48)$$

Typically, for an isotropic turbulence structure and a 10 km horizontal path at a 50 meter altitude

$$\Delta\phi = 0.005 \text{ radians}$$

*Assuming the process to be ergodic so that $\langle \Delta\theta^2 \rangle = \overline{\Delta\theta^2}$ where the bar indicates a time average.

For an initial polarization of $\phi = 45^{o}$,

$$\frac{\Delta\phi}{\phi} \sim 1 \text{ percent}$$

This result is consistent with measurements made by Goodwin[64] of Hughes Research Laboratories.

12.7 REFERENCES

1. Evans, A.; Brchyuski, M.P.; and Wacker, A.G. "The Radio Spectrum in Aerospace Communications," Technical Report ASD-TR-61-589; Volume IV, Absorption in Planetary Atmospheres and Sources of Noise, pp. 115-132; RCA Victor Company Ltd., Montreal Canada, August 1962.

2. Ko, H.C. "The Distribution of Cosmic Radio Background Radiation," Proceedings of the IRC, Vol. 46, No. 1, pp. 218-215, January 1958.

3. Brown, R., Hanbury and Hazard, C. "A Model of Radio-Frequency Radiation from the Galaxy," Philosophical Magazine, Volume 44, No. 7, pp. 939-936, September 1953.

4. Stephenson, R.G. "External Noise," Space Communications, Edited by A.V. Balikrishnan, McGraw-Hill Book Company, Inc.

5. Cottony, H.V. and Johler, J.R. "Cosmic Radio Noise Intensities in the UHF Band," Proceedings of the IRE, Volume 40, pp. 1487-1489, 1946.

6. Piddington, J.H. and Trent, G.H. "A Survey of Cosmic Radio Emission at 600 mc," Australian Journal of Physics, Volume 9, pp. 481-493, December 1956.

7. Smith, A.G. "Extraterrestrial Noise as a Factor in Space Communications," Proceedings of the IRE, Volume 48, No. 4, pp. 594, April 1960.

8. Stephenson, R.G. op. cit.

9. Ko, H.C. op. cit.

10. Conway, R.G.; Kellerman, K.I.; and Long, R.F., "The Radio Frequency Spectra of Discrete Radio Sources," Monthly Notices of the Royal Astronomical Society, Volume 125, pp. 268-269, 1963.

11. Ibid., pp. 261-284.

12. Smith, A.G., op. cit., p. 595.

13. Evans, Bochyuski, and Wacker, op. cit.

14. Hogg, D.C. and Mumford, W.W. "The Effective Noise Temperature of the Sky," Microwave Journal, Volume 3, pp. 80-84, March 1960.

15. Malitson, H.H. "The Solar Energy Spectrum," Sky and Telescope, pp. 162-165, March 1965.

16. Mayer, C.H. "Thermal Radio Radiation from the Moon and Planets," IEEE Transactions on Antennas and Propagation, Volume AP 12, No. 7, pp. 902-913, December 1964.

17. Mayer, C.H., McCullough, T.P., and Sloanaker, R.M. "Meaurements of Planetary Radiation at Centimeter Wavelengths," Proceedings of the IRE, Volume 46, No. 1, pp. 260-266, January 1958.

18. Smith, A.G., op. cit., p. 598.

19. Gibson, J.E. "Lunar Thermal Radiation at 35 kmc," Proceedings of the IRE, Volume 46, No. 1, pp. 280-286, January 1958.

20. Stephenson, R.G., op. cit., p. 98.

21. Millman, G.H. "Atmospheric Effects on VHF and UHF Propagation," Proceedings of the IRE, Volume 48, No. 8, pp. 1492-1501, August 1958.

22. Blake, L.V. "Tropospheric Absorption and Noise Temperature for a Standard Atmosphere," Summary of a Paper Presented at the 1963 PT-GAP International Symposium, NBS, Boulder Colorado, July 9-11, 1963.

23. Ramsey, R.C. "Spectral Irradiance from Stars and Planets, above the Atmosphere, from 0.1 to 100 microns," Applied Optics, Volume I, No. 4, July 1962.

24. McGurn, J.F. "Solar System Radiation," Hughes Aircraft Company IDC 2729.1/64, June 1966.

25. McGurn, J.F., Hughes Aircraft Company, private communcation, June 1966.

26. Kuiper, G.E. The Solar System, Volume III, Planets and Satellites, University of Chicago Press, 1961.

27. Ibid.

28. Goldberg, I.L. "Radiation from Planet Earth," U.S. Army Research and Development Laboratory, Fort Monmouth, New Jersey, September 1961.

29. Dennison, A.J. "Illumination of a Space Vehicle Surface Due to Sunlight Reflected from Earth," Journal of the American Rocket Society, Volume 32, pp. 635-637, April 1962.

30. Dennison, A.J., "Illumination of a Cell Surface in Space Due to Sunlight Reflected from Earth," Report No. T1S61SD101, General Electric Co., June 1961.

31. Olson, G.G. "Background Considerations for Ground-Based Terminal Defense Sensors," Hughes Aircraft Company IDC 2230.03/189, October 1965.

32. Mitra, S.K. "The Upper Atmosphere," Asiatic Society, Calcutta, India, 1962.

33. Roach, F.E. "Manual of Photometric Observations of the Airglow During the IGY," National Bureau of Standards Report No. 5006.

34. Olson, G.G., op. cit.

35. Geo-Science Inc. Final Report: "Twilight and Airflow Study," Air Force Contract AF19(122)-433, January 1953.

36. Ibid.

37. Ibid.

38. Millman, G.H., op. cit.

39. Blake, L.V., "Tropospheric Absorption and Noise Temperature for a Standard Atmosphere," Summary of a paper presented at the 1963 PT-GAP International Symposium, NBS Boulder, Colorado, July 1963.

40. Meyer, James W., "Radar Astronomy at Millimeter and Sub-millimeter Wavelengths," Proceedings of the IEEE, Volume 54, Number 4, p. 488, April 1966.

41. Booker, H.S. and Gordon, W.E., "A Theory of Radio Scattering in theTroposphere," Proceedings of the IRE, Volume 38, Number 4, pp. 401-412, April 1950.

42. Muchmore, R.B. and Wheelon, A.D., "Line-of-Sight Propagation Phenomena—I. Ray Treatment," Proceedings of the IRE, Volume 43, Number 10, pp. 1437-1449, October 1955.

43. Wheelon, A.D. and Muchmore, R.B., "Line-of-Sight Propagation Phenomena—II. Scattered Components," Proceedings of the IRE, Volume 43, Number 10, pp. 1450-1458 , October 1955.

44. Wheelon, A.D., "Near Field Corrections to Line of Sight Propagation," Proceedings of the IRE, Volume 43, Number 10, pp. 1459-1466, October 1955.

45. Wheelon, A.D., "Radio-Wave Scattering by Tropospheric Irregularities," Journal of Research of the NBS, Volume 63D, Number 2, pp. 205-233 (1959).

46. Herbstreit, J.W. and Thompson, M.C., "Measurements of the Phase of Radio Waves Received over Transmission Paths with Electrical Lengths Varying as a Result of Atmospheric Turbulence," Proceedings of the IRE, Volume 43, Number 10, pp. 1391-1401, October 1955.

47. Dearn, A.P. and Fannin, B.M., "Phase Difference Variations in 9,350-Megacycle Radio Signals Arriving at Spaced Antennas," Proceedings of the IRE, Volume 43, Number 10, pp. 1402-1404, October 1955.

48. Thompson, M.C. and Janes, H.B., "Measurements of Phase Stability over a Low-Level Tropospheric Path, Journal of Research of the NBS, Volume 63D, Number 1, pp. 45-51 (1950).

49. Norton, K.A., et al., An Experimental Study of Phase Variations in Line-of-Sight Microwave Transmissions, NBS Monograph 33, November 1961.

50. Decker, M.T. and Guirand, F.O., Tropospheric Radio Propagation on an Oversea Path in the Gulf of Mexico, NBS Report 8429, May 1964.

51. Smith, P.L., Scattering of Microwaves by Cloud Droplets, Proc. World Conference on Radio Meteorology, Boulder, Colorado, pp. 202-207, 1964.

52. Adler, J.E., et al., The Design of an Experiment to Determine the Limitations Imposed on a Multiple Aperture Antenna System by Propagation Phenomena, 3rd Quarterly Report, Contract No. NAS 5-3974, SRI Project 5067, Stanford Research Institute, Menlo Park, California, March 1965.

53. Chapman, R.M. and Carpenter, R., "Effect of Night Sky Backgrounds on Optical Measurements," Technical Report 61-23-A, Geophysics Corporation of America, May 1961.

54. Hogg, D.C., "Effect of the Troposphere on the Propagation of Coherent Optical Waves," Proceedings of the IEEE PT-GAP International Symposium, pp. 102-108, 1965.

55. Davis, J.I., "Consideration of Atmospheric Turbulence in Laser Systems Design," Applied Optics, Volume 5, Number 1, pp. 139-47, January 1966.

56. Davis, J.I., High Energy Laser Systems Analysis, Semi-annual Report, Hughes Aircraft Company Reference Number A 7910, 31 December 1965.

57. Hufnagel, R.E., "Understanding the Physics of Seeing Through Turbulent Atmospheres," abstracted from Advanced Range Instrumentation presentation made by Perkin Elmer Corporation at Patrick AFB, Florida, 20 November 1963.

58. Tatarski, V.I., Wave Propagation in a Turbulent Medium, McGraw-Hill Book Company, Inc., New York, 1961.

59. Davis, J.I., "Consideration of Atmospheric Turbulence in Laser Systems Design, "Applied Optics, Volume 5, Number 1, pp. 139-47, January 1966.

60. Tatarski, V.I., op. cit.

61. Davis, J.I., Hughes Aircraft Company, private communication, April 1966.

62. Ibid.

63. Hodara, H., "Laser Wave Propagation through the Atmosphere, Proceedings of the IEEE, Volume 54, Number 3, pp. 374, 375, March 1966.

64. Goodwin, F.E., Hughes Research Laboratories, private communication with H. Hodara.